

STIC-Biotech/ChemLib

138413

From: Chan, Christina  
Sent: Monday, November 22, 2004 10:30 AM  
To: Walicka, Malgorzata  
Cc: STIC-Biotech/ChemLib  
Subject: RE:

Please rush. Thanks Chris

Chris Chan  
TC 1600 New Hire Training Coordinator and SPE 1644  
(571)-272-0841  
Remsen, 3E89

-----Original Message-----

From: Walicka, Malgorzata  
Sent: Sunday, November 21, 2004 5:31 PM  
To: Chan, Christina  
Cc: STIC-Biotech/ChemLib  
Subject:

Please authorize the RUSH search of SEQ ID NO:4 in the application No. 10/802,682. Please also against pending data.

Thank you.

Malgorzata A. Walicka, Ph.D.  
Patent Examiner  
Art Unit 1652, Recombinant Enzymes  
USPTO, Remsen Building, Room 2C76  
400 Dulany St.  
Alexandria, VA 22313  
Mail Room 2C70  
Tel. (571)272-0944, fax (571)273-0944

RECEIVED  
NOV 22 2004  
STIC/BIOTECH DIVISION  
(STIC)

\*\*\*\*\*

STAFF USE ONLY

Searcher: Arnold  
Searcher Phone: 2-2532  
Date Searcher Picked up: 11/22/04  
Date Completed: 11/23/04  
Searcher Prep/Rev. Time:      
Online Time:    

\*\*\*\*\*

Type of Search

NA Sequence: # 1  
AA Sequence: #      
Structure: #      
Bibliographic:      
Litigation:      
Patent Family:      
Other:    

\*\*\*\*\*

Vendors and cost where applicable

STN:      
DIALOG:      
QUESTEL/ORBIS:      
LEXIS/NEXIS:      
SEQUENCE SYSTEM:      
WWW/Internet:      
Other(Specify):

**This Page Blank (uspio)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 2004, 13:16:30 ; Search time 7584 Seconds

(without alignments)  
10849.694 Million cell updates/sec

Title: US-10-802-682-4

Perfect score: 1740  
Sequence: 1 atgaaccccaacacgctgct.....tcttgactgcgcagctaa 1740

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenEmbl:.\*  
1: gb\_ba:.\*  
2: gb\_hhg:.\*  
3: gb\_in:.\*  
4: gb\_com:.\*  
5: gb\_ov:.\*  
6: gb\_pat:.\*  
7: gb\_ph:.\*  
8: gb\_pl:.\*  
9: gb\_pr:.\*  
10: gb\_ro:.\*  
11: gb\_sts:.\*  
12: gb\_sy:.\*  
13: gb\_un:.\*  
14: gb\_vi:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1740	100.0	1740	6	A70009 Sequence 4
2	1740	100.0	1740	6	CQ797619 Sequence 1
3	1130.2	65.0	1740	6	A70006 Sequence 1
4	1079.8	62.1	1737	6	A70008 Sequence 3
5	1053.6	60.6	1740	6	A70007 Sequence 2
6	625.8	36.0	22913	1	AB046580 Sequence 1
7	141.8	8.1	299910	1	AP005957 Bradyrhiz
8	118.2	6.8	299910	1	AP005957 Bradyrhiz
9	110.2	6.3	300600	1	AP005936 Bradyrhiz
10	110.2	6.3	2097	1	AF277373 Ralstonia
11	101.2	5.8	2959	1	CTOHEH X81880 C.lescoester
12	98.8	5.7	7097	1	AF176640 Pseudomon
13	98.4	5.7	302101	1	AE016784 Pseudomon
14	95.2	5.5	110000	1	AY316747 Rhizobium
15	94.8	5.4	2250	1	AF355798 Pseudomon
16	94.4	5.4	4534	1	AB091400 Pseudomon
17	92.4	5.3	300200	1	AP005962 Bradyrhiz
18	88.8	5.1	2214	6	E03223 The gene en
19	88.8	5.1	4230	1	ABCADHCC D00635 Acetobacter

20	87.2	5.0	2802	1	AEADHGENE	X82894 A.europaeus
21	87.2	5.0	4059	1	AEY09480	Y09480 A.europaeus
22	85.6	4.9	2163	6	AR143029	AR143029 Sequence
23	85.6	4.9	2163	6	BD269976	BD269976 Gluconoba
24	85.6	4.9	2265	6	AR143020	AR143020 Sequence
25	85.6	4.9	2265	6	BD269967	BD269967 Gluconoba
26	85.6	4.9	4830	6	AR143023	AR143023 Sequence
27	85.6	4.9	4830	6	BD269970	BD269970 Gluconoba
28	81.2	4.7	2272	1	AY048591	AY048591 Pseudomon
29	80.8	4.6	5602	1	D86375	D86375 Gluconobact
30	78	4.5	302101	1	AE016784	AE016784 Pseudomon
31	77.2	4.4	1694	1	PDECTT553I	W75583 Paracoccus
32	77.2	4.4	9470	1	PD034346	U34346 Paracoccus
33	77	4.4	4915	1	PAE9858	AU009588 Pseudomon
34	77	4.4	1533	1	AE004624	AE004624 Pseudomon
35	76.6	4.4	349635	1	BX572603	BX572603 Rhodospir
36	74.8	4.3	1869	1	AB162821	AB162821 Pseudomon
37	74	4.3	1989	1	AR326086	AR326086 Pseudomon
38	70.4	4.0	2625	6	AX079058	AX079058 Sequence
39	63	3.6	125020	9	AF429315	AF429315 Homo sapi
40	60.4	3.5	145153	1	AY458649	AY458649 Uncultured
41	60.2	3.5	2314	1	PDEM0H	M17339 P.denitrific
42	60.2	3.5	125020	9	AF429315	AF429315 Homo sapi
43	60	3.4	2011	1	AF309488	AF309488 Simorhizo
44	60	3.4	110000	1	RWE51985_01	Continuation (2 of
45	59.8	3.4	2466	1	MTBXXDH	M22625 Methylobact

#### ALIGNMENTS

RESULT 1	A70009	Sequence 4 from Patent EP0832974.	1740 bp	DNA	linear	PAT 07-MAY-1999
LOCUS	A70009	Sequence 4 from Patent EP0832974.				
DEFINITION	A70009	Sequence 4 from Patent EP0832974.				
ACCESSION	A70009	Sequence 4 from Patent EP0832974.				
VERSION	A70009.1	GI:4774465				
KEYWORDS						
SOURCE						
ORGANISM		unidentified				
		unclassified.				
REFERENCE		1 (bases 1 to 1740)				
AUTHORS		Tomiyama, N., Asakura, A., Ojima, S., Hoshino, T. and Shinjoh, M.				
TITLE		Alcohol-aldehyde-dehydrogenases				
JOURNAL		Parent: EP 0832974-A 4 01-Apr-1998;				
		HOFFMANN LA ROCHE (CH)				
FEATURES						
source		Location/Qualifiers				
		1..1740				
		/organism="unidentified"				
		/mol_type="unassigned DNA"				
		/db_xref="taxon:32644"				

#### ORIGIN

Query Match	100.0%	Score 1740;	DB 6;	Length 1740;
Best Local Similarity	100.0%	Pred. No. 6,1e-309;		
Matches 1740;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
1	ATGAACCCCAACGCTGCTTCGACACGCGCGCGGCTGCTATTGCTTACCGCGCCGCC	60		
1	ATGAACCCCAACGCTGCTTCGACACGCGCGCGGCTGCTATTGCTTACCGCGCCGCC	60		
1	ATGAACCCCAACGCTGCTTCGACACGCGCGCGGCTGCTATTGCTTACCGCGCCGCC	60		
61	GCATTTCGCGAGTAAACCCGATTAACGATGAACCTGCGGAAACCCCGCTGTGAA	120		
61	GCATTTCGCGAGTAAACCCGATTAACGATGAACCTGCGGAAACCCCGCTGTGAA	120		
121	TGATTAACGCGCGCGCAACCAAGAACTATGCGACCTGCGCCGACCGCATCTACT	180		
121	TGATTAACGCGCGCGCAACCAAGAACTATGCGACCTGCGCCGACCGCATCTACT	180		
181	GCCGCAACGTTGGTCACTGCTGCGCGCGCGCGGATGAGAGCGCGGCGCGCTA	240		
181	GCCGCAACGTTGGTCACTGCTGCGCGCGCGCGGATGAGAGCGCGGCGCGCTA	240		
181	GCCGCAACGTTGGTCACTGCTGCGCGCGCGCGGATGAGAGCGCGGCGCGCTA	240		

QY	241	CAGGTCAACGGCGATGATCAAGATGAGGGCTGATGATCTGGCAAAACCCCGGATGATGATC	300
Db	241	CAGGTCAACGGCGATGATCAAGATGAGGGCTGATGATCTGGCAAAACCCCGGATGATGATC	300
QY	301	CAGGCGCTGATGCGCAAAAGCGGATCTGATCTGGGAAACAACGCGCGCAACTGCCGCC	360
Db	301	CAGGCGCTGATGCGCAAAAGCGGATCTGATCTGGGAAACAACCGCGCGCAACTGCCGCC	360
QY	361	GTGGCGACGCTAAACGCCCAAGGCGGACCCGCAAGGCGCGCGCTTTCACGACAGC	420
Db	361	GTGGCGACGCTAAACGCCCAAGGCGGACCCGCAAGGCGCGCGCTTTCACGACAGC	420
QY	421	CTCTATTTCAGCTCATGAGGACAACTATCTGATCGCGCTGATATAGAGACGGGCAAGTC	480
Db	421	CTCTATTTCAGCTCATGAGGACAACTATCTGATCGCGCTGATATAGAGAGCGGCAAGTC	480
QY	481	GTAATTCATGTCGAACGTGATTCGGGCGGAAACGGCTTGACCACTAACACACGCGGGCCG	540
Db	481	GTAATTCATGTCGAACGTGATTCGGGCGGAAACGGCTTGACCACTAACACACGCGGGCCG	540
QY	541	ATTGTCGCAATGCGGCTCATGCTCGCGGGTTTCACTGCCAATATTGCGCCTATAGATGC	600
Db	541	ATTGTCGCAATGCGGCTCATGCTCGCGGGTTTCACTGCCAATATTGCGCCTATAGATGC	600
QY	601	TTTATCTCGGGCACAGATTCCGCGACGCGGTAGAGACTGTGGCCAACTTATATCCG	660
Db	601	TTTATCTCGGGCACAGATTCCGCGACGCGGTAGAGACTGTGGCCAACTTATATCCG	660
QY	661	CAGCGGGCGAAGAGGGTGAAGAGCTTGGGGCAATGATTTGAGAGCGCGCTGATGAC	720
Db	661	CAGCGGGCGAAGAGGGTGAAGAGCTTGGGGCAATGATTTGAGAGCGCGCTGATGAC	720
QY	721	GCGGCTGAGGGGTCAGATCACTATGATCCGCTGACGAACCTTGTGTTCTATGCTGAC	780
Db	721	GCGGCTGAGGGGTCAGATCACTATGATCCGCTGACGAACCTTGTGTTCTATGCTGAC	780
QY	781	GCGGTGGGGCCCAAGCTCCGAAACCAAGCGGGGACCGCGCGGGACGCTGTATGAGAAC	840
Db	781	GCGGTGGGGCCCAAGCTCCGAAACCAAGCGGGGACCGCGCGGGACGCTGTATGAGAAC	840
QY	841	AACACCCGCTTGGCGGTGCTGCCACACGCGGCGAGATTGTCTGCGCTCACAGACCTG	900
Db	841	AACACCCGCTTGGCGGTGCTGCCACACGCGGCGAGATTGTCTGCGCTCACAGACCTG	900
QY	901	CCGCGCGCAACTGTGGGACCAAGATGCACTTTCGATGATGATGATGATGATGATGATG	960
Db	901	CCGCGCGCAACTGTGGGACCAAGATGCACTTTCGATGATGATGATGATGATGATGATG	960
QY	961	CAACCTCGAGCGAGATGAGAGGATCTGCGGCCCACTAACCCCAATGCGGCGACCGGCGAG	1020
Db	961	CAACCTCGAGCGAGATGAGAGGATCTGCGGCCCACTAACCCCAATGCGGCGACCGGCGAG	1020
QY	1021	CGCGCTGTGCTGAACGGGTGCGCTTTCGAAAGCCCGGACGAGTGTGGCTTGTGATGCGGCC	1080
Db	1021	CGCGCTGTGCTGAACGGGTGCGCTTTCGAAAGCCCGGACGAGTGTGGCTTGTGATGCGGCC	1080
QY	1081	TGCGGCGAATTCCTGTGGGCGGCTGATACCACTACCACTATGATGATGATGATGATGATG	1140
Db	1081	TGCGGCGAATTCCTGTGGGCGGCTGATACCACTACCACTATGATGATGATGATGATGATG	1140
QY	1141	GAGACCGGCTTGTGACGCTGAACGAGATGCGGTGCTGAAAGACTGGAAGCTGGAATAT	1200
Db	1141	GAGACCGGCTTGTGACGCTGAACGAGATGCGGTGCTGAAAGACTGGAAGCTGGAATAT	1200
QY	1201	GACGCTGCGCGCACTTCTCGGTGGGTGGGCGACATGCTGTGACGCGCGCACTGAACCCGGAC	1260
Db	1201	GACGCTGCGCGCACTTCTCGGTGGGTGGGCGACATGCTGTGACGCGCGCACTGAACCCGGAC	1260
QY	1261	ACCGGCAATTTACTTTCTTTCGCTGGAACAATGCTCTCTACATATTATGACCGCTTGAACA	1320
Db	1261	ACCGGCAATTTACTTTCTTTCGCTGGAACAATGCTCTCTACATATTATGACCGCTTGAACA	1320
QY	1321	GAGTTTACGCGCTGACGCTCTATATAACACAGCGGACCGGAAACTGCGCGCGGCTTT	1380

Db	1321	GAGTTTAGCGCGCTCGACGTCCTATTAAACACGACGCCCGACCCGCAAAACTCGGCGCGGCGTTT	1380
Qy	1381	GAAATATATGCGCGCGCATTCGACGCGATTGATATCGACCGGCGCGACCTTGTGCTCGCG	1440
Db	1381	GAAATATATGCGCGCGCATTCGACGCGATTGATATCGACCGGCGCGACCTTGTGTGCGCG	1440
Qy	1441	GAGCGCCCTGCGCGGAACTACTCGCCCGTTTGTGACGCGCAGGCGGCTGTGTGTTCAAC	1500
Db	1441	GAGCGCCCTGCGCGGAACTACTCGCCCGTTTGTGACGCGCAGGCGGCTGTGTGTTCAAC	1500
Qy	1501	GGCGGACACGACCGGCTATTTTCCGCGCCCTACGACCGAGAAAACCGCGCAGACTTTGTGCGAG	1560
Db	1501	GGCGGACACGACCGGCTATTTTCCGCGCCCTACGACCGAGAAACCGCGCAGACTTTGTGCGAG	1560
Qy	1561	GCCCGCTTTGCGACGCGTTCGCGACCGGCGCAGGCGCATCGACTACGAGTTTGAACGCGCGTGCAA	1620
Db	1561	GCCCGCTTTGCGACGCGTTCGCGACCGGCGCAGGCGCATCGACTACGAGTTTGAACGCGCGTGCAA	1620
Qy	1621	TATATCGCATCGGTGCGGCGGCTGTGACCTATGCGACGAAATTGAACGCGCGCGCTGGCC	1680
Db	1621	TATATCGCATCGGTGCGGCGGCTGTGACCTATGCGACGAAATTGAACGCGCGCGCTGGCC	1680
Qy	1681	GAGCGAATCGATTGACCTCGGTGCGGTAAATGCGATCTATGCTTTTGGACATGCGCGCAGTAA	1740
Db	1681	GAGCGAATCGATTGACCTCGGTGCGGTAAATGCGATCTATGCTTTTGGACATGCGCGCAGTAA	1740

RESULT 2					
CQ97619					
LOCUS					
DEFINITION	CQ97619	1740 bp	DNA	linear	PAT 20-APR-2004
ACCESSION	Sequence 1	From Patent	WO2004029267.		
VERSION	CQ97619				
KEYWORDS	CQ97619.1	GI:4642899			
SOURCE					
	Gluconobacter oxydans				

**ORGANISM** *Gluconobacter oxydans*  
Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;  
Acetobacteraceae; *Gluconobacter*.

REFERENCE 1  
AUTHORS Hoshino, T. and Shinjoh, M.  
TITLE Process for producing L-ascorbic acid  
JOURNAL Patent: WO 2004029267-A 1 08-APR-2004;

## FEATURES

### Location/Qualifiers

```

/organism="Glucobacter oxydans"
/mol_type="unassigned DNA"
/db_xref="taxon:442"
1..1740
CDS

```

CDS

```

/note="unnamed protein product"
/codon_start=1
/translation_table=1
/protein_id="CAG36844.1"
/db_xref="GI:46425900"
/translation="MNPPTLLRTAAVLLTPAPAFQVPTIDELILANPAGEMINV
GRNENYRHSPLTQITADNVQQLDVMARGEAAAVOTPMIHGVNLANPGVIOQ
LDAQTGDLIWEHRQLPAVATLNAQDPKRRVALYGSILYESNDNHIALDMETGO
VFEDVRGSGEDGLTSNTGPIVANGVYVAGSTQYSPYGCISGHDASITGELRNHNE
IPQGEEDDETWDGDFEARMMYGVAGTIVPVNLVFGSGTQPGASLETQGTGPGG
LYGNTRPRAVAPDPTGELVYMRQTLPRNMWDECFEEMVNAVDPQPSMEGLPALIN
NAAGERRVLTPGAPCKTGITWMSFDAASGEFLIMADNTYTNMIAISIDEGLYTVNEDA
LKEHLEDVYDQPTFLGGRDMSAALNPDTGYFLPLNNACVDINAVQEFSAIVNTN
SATKLAFENMGRILDAIDISTGRTLWASLRPAANYSPVLTSGGVFENGTRIVFYF
ALSQETGTLTMQARLATATATQGLISTELDWQYVIALAGGULTYGTQNALPAAEIDS
SVGNALIVFALPQ"

```

## ORIGIN

Query Match	100.0%	Score 1740;	DB 6;	Length 1740;
Best Local Similarity	100.0%	Pred. No. 6.1e-309;		
Matches 1740; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

1 ATGACCCCAACAGCTGCTTGGACCAAGCGCGCCGTGTAATTGCTTACCGCGCCGCC 60

```

Db      1 ATGAACCCCAACAGCTGCTTCGACACGCGCGCTGCTATTTGCTAACCGGCCGCC 60
Qy      61 GCATTCGCGAGAGTAACCCCGATTAACCGATGAAGTGTGTGCGAAACCGCGCGCTGTGAA 120
Db      61 GCATTCGCGAGAGTAACCCCGATTAACCGATGAAGTGTGTGCGAAACCGCGCGCTGTGAA 120
Qy      121 TGAATTAACAGGCGCCCAACCAAGAAATCATTCGCACTCGCGCCGTGAACCGATCACT 180
Db      121 TGAATTAACAGGCGCCCAACCAAGAAATCATTCGCACTCGCGCCGTGAACCGATCACT 180
Qy      181 GCCGACAACTTGTGCTGATTGCAACTGTGTGCGCGCGCGAGTGAAGCGGCGCGCTGA 240
Db      181 GCCGACAACTTGTGCTGATTGCAACTGTGTGCGCGCGCGAGTGAAGCGGCGCGCTGA 240
Qy      241 CAGGTACGCGCGAGTATCATATGAGTGTATTCGCAAAACCCCGGTGATGATGATC 300
Db      241 CAGGTACGCGCGAGTATCATATGAGTGTATTCGCAAAACCCCGGTGATGATGATC 300
Qy      301 CAGGCGCTGATGCGCAACAGGCGATCTGATCTGGGAAACACCGCGCGCAACTGCGCGC 360
Db      301 CAGGCGCTGATGCGCAACAGGCGATCTGATCTGGGAAACACCGCGCGCAACTGCGCGC 360
Qy      361 GTGCGCACTGTAACCGCGCAAGCGGCGACCGGACCGCGCGCTGCGCTTTACGCAAGAC 420
Db      361 GTGCGCACTGTAACCGCGCAAGCGGCGACCGGACCGCGCGCTGCGCTTTACGCAAGAC 420
Qy      421 CTCTATTTACGCTCATATGAGCAACCATCTGATCGCGCTGATATGAGAGCGGCGCAGT 480
Db      421 CTCTATTTACGCTCATATGAGCAACCATCTGATCGCGCTGATATGAGAGCGGCGCAGT 480
Qy      481 GTATTCGATGTCGAAGCTGATCGGCGGAAAGACGCTTGAACGTAACACACGCGGCGC 540
Db      481 GTATTCGATGTCGAAGCTGATCGGCGGAAAGACGCTTGAACGTAACACACGCGGCGC 540
Qy      541 ATGTGCGCAATGAGCGTATCGTGCAGGCGTTCACCGCGCAATATTCGCGCTATGATG 600
Db      541 ATGTGCGCAATGAGCGTATCGTGCAGGCGTTCACCGCGCAATATTCGCGCTATGATG 600
Qy      601 TTTATCTCGGGGACGATTCGCGACGCGGTGAGAGCTGTGCGCAACACTTTATCCG 660
Db      601 TTTATCTCGGGGACGATTCGCGACGCGGTGAGAGCTGTGCGCAACACTTTATCCG 660
Qy      661 CAGCGCGGGAAGAGGTGACGAGCTTGCGCGCAATGATTTGAGGCGCGCTGATGAC 720
Db      661 CAGCGCGGGAAGAGGTGACGAGCTTGCGCGCAATGATTTGAGGCGCGCTGATGAC 720
Qy      721 GCGGCTGCGGGTCAATCACTATGATCCCGTGAACGACTTGTGTTCTATGCGTCAAC 780
Db      721 GCGGCTGCGGGTCAATCACTATGATCCCGTGAACGACTTGTGTTCTATGCGTCAAC 780
Qy      781 GCGGTGCGGCGCAGCGTTCGAAACCCAGCGCGGCGACCGCGCGCGCTGATGCGAC 840
Db      781 GCGGTGCGGCGCAGCGTTCGAAACCCAGCGCGGCGACCGCGCGCGCTGATGCGAC 840
Qy      841 AACACCCGCTTGTGCGGTGCGTCCGACACGCGGCGAGATGTCTGCGGTGACGACCGCT 900
Db      841 AACACCCGCTTGTGCGGTGCGTCCGACACGCGGCGAGATGTCTGCGGTGACGACCGCT 900
Qy      901 CCGCGGGAACAATGAGAGGTCTGCGCGCATGAACCCCAATGCGCGCGACGCGCGAG 960
Db      901 CCGCGGGAACAATGAGAGGTCTGCGCGCATGAACCCCAATGCGCGCGACGCGCGAG 960
Qy      961 CAACCTCGGCGAGATGAGAGGTCTGCGCGCATGAACCCCAATGCGCGCGACGCGCGAG 1020
Db      961 CAACCTCGGCGAGATGAGAGGTCTGCGCGCATGAACCCCAATGCGCGCGACGCGCGAG 1020
Qy      1021 CGCGGTGCTGAGCGGCTGCGCTTGAAGACGCGGCGAGATGCGTGTGATGCGGCGC 1080
Db      1021 CGCGGTGCTGAGCGGCTGCGCTTGAAGACGCGGCGAGATGCGTGTGATGCGGCGC 1080
Qy      1081 TCGGCGGAATTCCTGTGCGCGGTGATACCAATATGATGCGCTGATGAC 1140

```

```

Db      1081 TCGGCGGAATTCCTGTGCGCGGTGATACCAATATGATGCGCTGATGAC 1140
Qy      1141 GAGACCGGCTTTGTGACGCTGAACGAGATGCGGTGCTGAAGAGCTGAGCTGGAATAT 1200
Db      1141 GAGACCGGCTTTGTGACGCTGAACGAGATGCGGTGCTGAAGAGCTGAGCTGGAATAT 1200
Qy      1201 GACGTCTCGCGACCTTCTGCGGTGCGGCGAGCTGCTGATGAGCGGCACTGAACCGGAC 1260
Db      1201 GACGTCTCGCGACCTTCTGCGGTGCGGCGAGCTGCTGATGAGCGGCACTGAACCGGAC 1260
Qy      1261 ACCGCGATTTACTTCTTCCGCTGAACATGCTGCTGATATTAATGCGCGTGAACA 1320
Db      1261 ACCGCGATTTACTTCTTCCGCTGAACATGCTGCTGATATTAATGCGCGTGAACA 1320
Qy      1321 GAGTTAAGCGCGCTGAGCTATTAACACGAGCGGCGACCGGAAACTGCGCGCGCTTT 1380
Db      1321 GAGTTAAGCGCGCTGAGCTATTAACACGAGCGGCGACCGGAAACTGCGCGCGCTTT 1380
Qy      1381 GAAATATGAGCGCGATGACGCGATGATATGACACCGGCGCACTTGTGCGCGC 1440
Db      1381 GAAATATGAGCGCGATGACGCGATGATATGACACCGGCGCACTTGTGCGCGC 1440
Qy      1441 GAGCGCCCTGCGGCGAACTACTCGCCGCTTTGTGACAGCGGCGGTGTGTTCAAC 1500
Db      1441 GAGCGCCCTGCGGCGAACTACTCGCCGCTTTGTGACAGCGGCGGTGTGTTCAAC 1500
Qy      1501 GCGCGGACCGACCGCTATTTCTGCGCTCAGCAGGAAACCGGCGAGACTTTGTGCGAG 1560
Db      1501 GCGCGGACCGACCGCTATTTCTGCGCTCAGCAGGAAACCGGCGAGACTTTGTGCGAG 1560
Qy      1561 GCGCGCTTGTGCGAGCGGTGCGACGCGGCGATCACTATGACGAGTTGAGCGGCGTCAA 1620
Db      1561 GCGCGCTTGTGCGAGCGGTGCGACGCGGCGATCACTATGACGAGTTGAGCGGCGTCAA 1620
Qy      1621 TATATGCGATCGGCGGCGGCGGTCTGACCTATGAGCGCAATGAAACGCGCGCTGCGC 1680
Db      1621 TATATGCGATCGGCGGCGGCGGTCTGACCTATGAGCGCAATGAAACGCGCGCTGCGC 1680
Qy      1681 GAGGCAATGATTCGACCTCGGTGCGGTATGATGATGCTATGCTTTGACACTGCGCAGTAA 1740
Db      1681 GAGGCAATGATTCGACCTCGGTGCGGTATGATGATGCTATGCTTTGACACTGCGCAGTAA 1740

RESULT 3
LOCUS       A70006               1740 bp      DNA      linear      PAT 07-MAY-1999
DEFINITION  Sequence 1 from Patent EP0832974.
ACCESSION   A70006
VERSION     A70006.1 GI:4774462
KEYWORDS
SOURCE      unidentified
            unclassified.
ORGANISM    unidentified.
REFERENCE   1 (bases 1 to 1740)
AUTHORS     Tomiyama,N., Asakura,A., Ojima,S., Hoshino,T. and Shinjoh,M.
TITLE       Alcohol-aldehyde-dehydrogenases
JOURNAL     Hoffmann LA ROCHE (CH)
            Location/Qualifiers
            source
              1..1740
              /organism="unidentified"
              /mol_type="unassigned DNA"
              /db_xref="taxon:32644"

ORIGIN
Query Match      65.0%; Score 1130.2; DB 6; Length 1740;
Best Local Similarity 78.8%; Pred. No. 3.6e-197;
Matches 1374; Conservative 0; Mismatches 363; Indels 6; Gaps 2;

Qy      1 ATGAACCCCAACAGCTGCTTCGACACGCGCGCTGCTATTTGCTAACCGGCCGCC 60
Db      1 ATGAACCCCAACAGCTGCTTCGACACGCGCGCGCTGCTATTTGCTAACCGGCCGCC 60

```

QY	61	GCATTCGGGCAAGGTTAAACCCCGATTACCGATGAACCTGTGGGGAACCCGCGCTGGTGAA	120
Db	61	GCTTTGCTCAAGTGAACCCCGCTCACCGATGAATTGTCTGGGGAACCCGCGCTGGTGAA	120
QY	121	TGGATTAACTACGCGCGCAACCCAGAAAACTATGCGCACTTGCGCCCTGACCCAAATCACT	180
Db	121	TGGATCACTACGCGTCAAGAACCCAGAAAACTAACCGTCACTCGCCCTGACCCAAATCAAG	180
QY	181	GCCGCAACAGTTGGTCAATTGGCAACTGGTCTGGGCGCGCGGGATGAGAGCGGGGGCGTA	240
Db	181	ACTGGAACGTTGGGCAACTGTCAACTGGTCTGGGCGCGCGGCAATGACGCCGGCAAAAGTC	240
QY	241	CAGGTACAGCCGATGATCATGATGGCGTGATGTACTTGGCAAAACCCCGTGATGTATC	300
Db	241	CAAGTACAGCCCGCTGATCATGACGGCTGATGTACTTGGCAAAACCCCGGCGAAGTGTATC	300
QY	301	CAGGCGCTGATGAGCCCAACAGCGGCAATGATCTGAGGAACACCGCGCGCAACTCGCGCC	360
Db	301	CAGGCGCATGACGCCCAAACTGGCGCATGTGATCTGGGAACACCGCGCGCAACTGCGCAAC	360
QY	361	GTGCGACAGCTAAACGCCCAAGGCGACCGCAAGCGCGCGGTGCGCCTTTACGGACGAGC	420
Db	361	ATGCGACAGCTGAACAGGCTTTGGCGAGCGAACCGCGCGCATGCGCTTACGGACCAAC	420
QY	421	CTCTATTTCAGCTCATGAGAACCAACATCTGATCGGCTGGATATGAGAACCGGGCGAAGTC	480
Db	421	GTTTACTTTGTTTCGTGGGAACAACCACTGGTGGCGCTTGACACCGCAACTGGCCCAAGTC	480
QY	481	GTATTCGATGTGCAACGTGAGATCGGCGCAAGACGGCTTGACACGATTAACACACAGGGCGG	540
Db	481	ACGTTGCAAGTGAACCGCGGCCAACGGCAAGACAGTGTGG---ACTCGTGGGCGCG	537
QY	541	ATTGTGCGCAATGCGGTCATGTCGCGGGTTTCACCTGCGCAATATTGCGCTATGAGATGC	600
Db	538	ATCGTGGCAAAAGCGGTGATCGTTGCCGGTTGCACTTCGCAATATCTCCCGTGGCGTCG	597
QY	601	TTTATCTGGGGGCAAGATTCCGCGAGCGGGTGAAGCTGTGGCGCAACCACTTATCCCG	660
Db	598	TTTGTCTGGGGCAAGACTCGGCCACCGGTGAAGGCTGTGGCGCAACTTACTTATCCCG	657
QY	661	CAGCGGCGGAAGAGGGTGAACGAGACTTGGGGCAATGATTTGAGGCGCGCTGATGAAC	720
Db	658	CCGCGTGGGAGAGGGGTATGAGACTTGGGGCAACGATTAAGAAACCGTGTGATGAATGC	717
QY	721	GCGGTCTGGGGTCAATATCACTATGATCCCGTGAAGAACTTTGTCTATGGGCTGACG	780
Db	718	GGTGGCTGGGGCAATCACTATGACCCCGTCAACAACCTTTGTCACTACGCTTCGAC	777
QY	781	GCGTGGGCGCAAGCGTCCGAAACCCAGCGCGGCAAGCGGCGGCAAGCTGTATGGCAC	840
Db	778	GCTGTGGGTCCGCGGTCCGAAACCAACGCGGCAACCCGGGGGGCAGCGTGAACGGCACG	837
QY	841	AAACCCGCTTTGCGGTGCTGCCGACACGGGCGAGATTGTCTGGCGTCACTAGCCCTG	900
Db	838	AAACCCCGTTTGGCGGTGCTGCTTACACAGGCGAGATTGTCTGGCGTCACTAGCCCTG	897
QY	901	CGCGCGCAACCTGGGACCAAGATGACAGTTCCGAGATGATGCTGGCAACGTCGATGTG	960
Db	898	CCCCCGAACAATTGGGACCAAGAAATGCAAGTTCCGAGATGATGTCAACGAATGTGATTC	957
QY	961	CAACCTCGCGGAGATGAGAGGCTTGGCGGCTCAATMAACCCCAATGCGGCGAGCGGAG	1020
Db	958	CAACCTCGACCGAGATGAGAGGCTTGGCGGCTCAATMAACCCCAAGCGGCAACTGGCGAG	1017
QY	1021	CGCCGTGCTCAACGCGGTGGCGCTTGCAAGACCGGACGATGTGGTCTGTTATCGCGCC	1080
Db	1018	CGTGGGTCTGACCGCGCTTCCGTGCAAAACCGGCAACCATGTGGCAATTCGACCGGAA	1077
QY	1081	TCGGGGAATTCCTGTGGCGCGGTGATTAACAACTACCAATATGATGCGCTGATGAC	1140
Db	1078	ACCGGGGAATTCCTGTGGCGCGGTGATTAACAACTACCAAGATGATGCAATTCATCAC	1137
QY	1141	GAGACCGGCGCTTGTACGCTGAACAGAGATGGCGTCTGAAGAGCTGACGTTGAATAT	1200

Db	1138	GAAAAACGACATCGAGACCCGTGAAGGAAGATGCGCATCTCTGAAGAAACATGTGTAATAT	1197
QY	1201	GACGCTGCGCCGAGACCTTCTCGGGGTGGGGGCGAGCATGTGTGTCAGCCGCGACTGAACCCGAGC	1260
Db	1198	GACGCTGTGCCGACCTTCTTGTGGGGGCGCGGACCTGGGCGGTGGCGGCACTGAACCCGAGC	1257
QY	1261	ACCGGCAATTTACTTCTTGGCGGCTGAACAAATBCTGTCTACGATATTATANGCCGCTGATCA	1320
Db	1258	AGCGGCAATCACTACTTCATCCGCGCTGAACAAGCTGCGATATAGACAAAGAGGCGCTGTGATCAG	1317
QY	1231	GAGTTTAGCGGCGCTCGACGTCATATACACCAACGCGACCCGAAACCTCGGCGCGGCTTT	1380
Db	1318	GAAATTCACCTGTAGTAGAGCTTATTAACAACAGCAAGTGAACAAAGCTGCGGCCGCGGCAAG	1377
QY	1381	GAAATAATGGCGCGGCACTGACGCGGATGATATACACACCGGGGCGACCTTGTGTGCGCG	1440
Db	1378	GATATGATCGGTCGTATTGACGCGATGACATCGACACGCGGTCTCTAAGCTGTGTGTCGAT	1437
QY	1441	GAGGCGCCCTGGGGGGAATCTACTGCGCCGTTTGTGCGACGGGAGGCGGNTGTGTTCAAC	1500
Db	1438	GAAAGTGTGTGGGGGAATATTGCGCCCGTCTTGTGCAACCGGCGGCGGCTTGTGTCAAC	1497
QY	1501	GCGCGGACCGGACCGCTATTTTCCGTGACCTCAGCCAGGAAACCGGCGAGACTTGTGTGCGAG	1560
Db	1498	GGTGGTACGAGTCGTTACTTCCGCGCCCTCAGCCAAAGAAACCGGCGAGACCTCTGTGCGAG	1557
QY	1561	GCCCGCTCTTGGGAGCGGTGCGGACGGGGGACGGGCAATCAAGCATGAGTTGGAACGGGCTGCAA	1620
Db	1558	ACCGGCGCTTGGAAACCGGTGCGGTGCGGGCGAGGCCAATCTCTTACGAGTTGAGCGCATGCAA	1617
QY	1621	TATATCGCATCGGTGCGGGGCGGTCTGACCTATNGGACGCAATTGAAACGCGCGCGCTGGC-	1679
Db	1618	TATGTGTGCGCATGCGAGGAGGTGGTGAAGCTATGAGCTCGGGCGTGAACCTCGGCACTGGCT	1677
QY	1680	--CGAGGCAATGCAATTGCACTCTCGGTGCGTAAATGCGAATTATGTCTTGTGCACTGCGCGAG	1737
Db	1678	GCGGAGCGAGTGCAGCTGCACCGCAATCGGTAAACCGCGTCTAAGTTCGCGCCTGCGCGAA	1737
QY	1738	TTA 1740	
Db	1738	TTA 1740	

[illegible]

QY	6	GCATTCGCGAGGATACCCCAATTACCGATGAACTGCTGGGCAACCCGCGCTGGTGA	120
Db	61	GCCTTGTCCCA---AACCCCATCAACGATGAAATGCTGGGGAACCCGCGCTGTGA	117
QY	121	TGGATTAACTACGCGCGCAACCAAGAAACTATGCGCACTGCGCCCTGACCCAGTACT	180
Db	118	TGGATCAACTACGCTCAGAACCAAGAACTACCGCCACTGCGCCCTGACGCGATTACC	177
QY	181	GCGCAACAAGTTGGTCAATTGCACTGGCTCGGCGCGCGGAGATGAGAGGGGGGCGTA	240
Db	178	GCAGACAAAGTGGGCCCACTGCACATTGGTCTGGCGCGCGGTATGGAAGCGGCAAGTC	237
QY	241	CAGGTCAAGCCGATGATCCATGATGGCGTGAATGTAATCTGGCAAAACCCCGTGAATG	300
Db	238	CAAGTAAACCCCGTTGTCATGACGCGCTCATATGATCTGGGAAACCCCGGTGACGTGATC	297
QY	301	CAGGCGCTGATGCGCAACAGGCGCATCTGATCTGGGAAACCGCGCGCACTGCGCGGC	360
Db	298	CAGGCGCATGAGCGCGCGCAACCGCGCATCTGATCTGGGAAACCGCGCGCACTGCGGAC	357
QY	361	GTCGCCACGCTAAAGCCCAAGGCGCACCGCAAGCGCGCGTGCCTTATCGGCACAGC	420
Db	358	ATGGCCACCGCTAAACAGCTTGGTGAAGCGCACCGCGCATGGCGCCTCTATGCGCAACAC	417
QY	421	CTCTATTTAGCTATGCGGACAAACCATCTGATCGCGCTGGATATGAGAAGCGGCGCAGTTC	480
Db	418	GTCATATTTGGTCTCGTGGGACAAACCACTGGTTCGCGCTGACACCTCGACCGGCAAGTC	477
QY	481	GTATTTGATGTGAACGTGGATCGGGCGAAGCGCTTGAACAGTAAACACACCGGGCCG	540
Db	478	GTAATTTGAGTGAATCGCGCTCAAGGCAACGATA---TGGCTTCGAACCTGTCGCGCGG	534
QY	541	ATTGTGCGCAATGAGCGTCACTCGTCGCGGTTTCCACTCGCCAAATTTGGCCTTATGATGC	600
Db	535	ATTGTGCGCAATGGGCTCATGTTGGGGCTCGACTGTGAGTAATTCGCGTTGCGCTGT	594
QY	601	TTTATCTCGGGGACGATTCGCGCAAGGCTGAGAGCTGTGGGCGAACCACTTATCCG	660
Db	595	TTGCTTTTGGGCGCAAGACTCGGCCACCGGTGAAGAGTGTGGGCGAACCACTTATCCG	654
QY	661	CAGCGCGGAGAGAGGTTGACGAGCTGGGGCAATATTTCAGAGCGGCTGAGTAC	720
Db	655	CGCGCGGAGAGAGGTTGATGAGACCTGGGCAATATTACAGAGCGGCTGAGTAC	714
QY	721	GCGGCTGGGGTCAAGTCACTATGATCCGTGACGAACCTTGTGTTCTATGGCTGACC	780
Db	715	GCGGTTTGGGCGCAATCACTATGACCCCGTGGCGCCTTGTGCACTACCGCACCTCA	774
QY	781	GCGGTGGGCGCAGCGTCCGAAACCAAGCGCGGACACGCGGCGGACGCTGTATGGAC	840
Db	775	GCAGTTGGCGCTGGGCGGAGATTCAGGCGGCAACCGTGTGGCGCTGATGATGGAC	834
QY	841	AACACCCGCTTTGGCGTGGCTCCGACACAGGCGCAATGTCTGGGCTCACCAACCTG	900
Db	835	AACACCCGCTTTGCTGTCCGCCCGGAGACCGGCGAGATGCTGGGGTCAACAACTCG	894
QY	901	CGCGCGCAACTGGGACCAAGAAATGCACTTTCAGATGATGGTTCGCAACGTCGATGTG	960
Db	895	CCCGCGCAACTGGGACCAAGAGTATCGTTGAGATGATGGTTCGTAACGTCGACGTTC	954
QY	961	CAACCTCGGCGGAAATGAGAGGTTCTGGCGGCTCAACCCCAATGCGCGACAGGCGAG	1020
Db	955	CAGCCTCGGCTGAATGAAAGGCTGACGCGCACTCAACCCGATCGCGCACAGGCGAG	1014
QY	1021	CGCGGTGTGCTACCGGATGCGCTTTCAGAGACCGGACGATGTGTCTGTTATGCGGC	1080
Db	1015	CGTGGGTGTGACCGCGGTTCCGCGCAAGAACGCGACATGTGGAGTTTCGACGCGGA	1074
QY	1081	TGCGGAGAAATTCGTTGGGCGCGGTGATACCACTACCAATATGATGCGCTTCGATGCAC	1140
Db	1075	ACCGGAGAAATTCGTGGGCGGCGACACAGCTATCAGAACTGATTCGAAAGCGTGAT	1134

Qy		1141	GAGACCCGCGCTTGAAGCGTGAAACGAGAGATCCGGTGTCTGAAAAGACTGCAGCTTTAAT	1200				
Db		1135	CCCGATGGTCGTGGATCATGTGAAGAGAATCTGGTCTGACCGAGCTGGAAAGTGGCTTAT	1194				
Qy		1201	GACGTCGCCCGAACCTTCTCCTGGGTGGGCGGCACTGTGTCTCAGCCGCACTGAACCCGGAC	1260				
Db		1195	GAAATCTGCCCGAACCTTCTCCTGGGTGGGCGGCACTGTGTCTCAGCCGCACTGAACCCGGAT	1254				
Qy		1261	ACCGGCACTTAATTCTTCTGGCGGCTGGAACAATAGCCCTGCTACGATATPATAAGGCGGTGATCAA	1320				
Db		1255	ACTGGCATCTATTTCATCCCGCTTGAAACAAGCCTCTTAGCCGATGAGCGGCTGTGACAA	1314				
Qy		1321	GAGTTTAGCGCGCTCGACGCTCTATAACACCAGGCGCACCGCAAACTCGCGCCGGCTTT	1380				
Db		1315	GAGTTAGCTCGGCTCGATGTGTATTAAGTCAGCGCTGACGCTATATACTGTGCCCCGGTTGG	1374				
Qy		1381	GAAAAATTTGGGCGCGCATTCGACCGGATGTGAATATAGCACCCGGGGCACCTTGTGTGGGCG	1440				
Db		1375	GAAAAATTTGGGCGCGGTATCGACCGCATATGACATAGACAACCGGCGCACGCTGTGTGGGCT	1434				
Qy		1441	GAGCGCCCTCGGCGGAACCTACTGACCCCGTTTTTGTGCAAGCGCAGGCGGTGTGTTCAAC	1500				
Db		1435	GAAAGCTAACGCTCTGAACTACGCGCTGTCTGTCTGTCCACCGGGCGGCGGTGTCTTCAAC	1494				
Qy		1501	GGGGGGAACCGACCGCTATTTCCGTGCCCCTCAGCCAGAGAAACGGGCGAGACTTTGTGGCAG	1560				
Db		1495	GGGGGGAACCGACCGCTTACTTCCGCGCCCTCAGCCAGAGAAACGGGCGAGAGCTGTGGCAG	1554				
Qy		1561	GCCCGTTTTCGACGCGGTGCGGACGGGGCAGGCGATGACGTACAGTTCGACCGCGGTGCAA	1620				
Db		1555	ACCGGTTTTCGACGCGGTGCGGACGGGGCAGGCGATTCCTATAGATTCGACCGCGGTGCAA	1614				
Qy		1621	TATATGCCATTCGATGCGGCGGCGGTCTGACCTATAGGACGCAATTGAAACGGCGCGCTGACC	1680				
Db		1615	TATGTGCGCATTCGGGCGGCGGCGGCAACAGCTATAGGACGCAACCAACCGCGGCTTGACC	1674				
Qy		1681	GAGGCAATGCATTTGCACCTTGGGTTCGGTAATGCGATCTATGTCTTTTGACATGCGCGAGTA	1739				
Db		1675	GAGCGGATCGACCTCGAACCGCCATCGGCAAGCGCATCTATGTCTTTGCTGCGCGAGCA	1733				
<b>RESULT 5</b>								
A70007								
LOCUS	A70007	1740 bp	DNA	linear PAT 07-MAY-1999				
DEFINITION	Sequence 2 from Patent EP0832974.							
ACCESSION	A70007							
VERSION	A70007.1	GI:4774463						
KEYWORDS	unidentified unclassified							
SOURCE	unclassified							
ORGANISM	unclassified							
REFERENCE	1 (bases 1 to 1740)							
AUTHORS	Tomiyama,N., Asakura,A., Ojima,S., Hoshino,T. and Shinjoh,M.							
TITLE	Alcohol-aldehyde-dehydrogenases							
JOURNAL	Patent: EP 0832974-A 2 01-APR-1998;							
	HOFFMANN LA ROCHE (CH)							
FEATURES	Location/Qualifiers							
SOURCE	1..1740							
	/organism="unidentified"							
	/mol_type="unassigned DNA"							
	/db_xref="taxon:32644"							
<b>ORIGIN</b>								
Query Match 60.6%; Score 1053.6; DB 6; Length 1740;								
Best Local Similarity 75.7%; Pred. No. 3,9e-183;								
Matches 1318; Conservative 0; Mismatches 419; Indels 3; Gaps 1;								
Qy	1	ATGAACCCCAACACGCTGCTTGACACGAGCGGCGCGGTATATGCTTACCGGCGCGGCC	60					
Db	1	ATGAAGACGTCGCTTGTGCTGTGGAGCGGTGGCGGCTTGCAAGCTATAGCTCTTT	60					
Qy	61	GCAATCGCGAGGTAAACCCGATTAACGATGAATCTGTCGACCAACCGGCGCGGTGTGA	120					

Db 61 GGGCTTGCTCAAGTAGACCCCGCTCAACCGATGATTTGCTGGGCAACCCCGCGCTGGTAA 120  
 Qy 121 TGGATTAACTACGCGCGCAACCAAGAAACTATCGCACTCGCCCTTGAACCAATCACT 180  
 Db 121 TGGATTACGCTACGCTAGAGAACCAAGAAACTATCGCACTCGCCCTTGAACCAATCACT 180  
 Qy 181 GCGCAACAAGTTGGTCAATTTGCACTGGTCTGGGCGCGGAGATGAGAGGCGGCGCTA 240  
 Db 181 ACTGAAAGATCGGCGCACTGCACTGGTCTGGGCGCGGAGATGAGAGGCGGCGCTA 240  
 Qy 241 CAGGTACCGCGATGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
 Db 241 CAGGTACCGCGATGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
 Qy 301 CAGGCGCTGGATGCGCAACAGCGGATGATGATGATGATGATGATGATGATGATGATGAT 360  
 Db 301 CAGGCGCTGGATGCGCAACAGCGGATGATGATGATGATGATGATGATGATGATGATGAT 360  
 Qy 361 GTGCGACGCTAAACCGCCCAAGCGCAACCGCGGCGCTGGCTTTACGCGACGAGC 420  
 Db 361 ATGCGACGCTGAAGAGCTTTGGCGAGCGGACCGCGGCGCTGATGAGCGACCAAC 420  
 Qy 421 CTCTATTTCAGCTCAGTGGCAACCATCTGATCGCGCTGATGATGATGATGATGATGATGAT 480  
 Db 421 GTTTACTTTTGTTCGTGGCAACCAACCTGGTCCCTGACACCGCACTGGCCAAAGT 480  
 Qy 481 GTATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540  
 Db 481 AGCTTGACGCTGACCGCGCGCAAGCGGCAAGACATGCTTTC---AAGCTGCGGCGCG 537  
 Qy 541 ATTGTGCAATGCGCTCATCTGCGGCTTCACTGCGCAATTTGCGCTCATGATGATGATGAT 600  
 Db 541 ATTGTGCAATGCGCTCATCTGCGGCTTCACTGCGCAATTTGCGCTCATGATGATGATGAT 600  
 Qy 538 ATGTGTGCAAAACGCGCTGATCTGCTGCGGCTGACCTGCAATCTGCGCTGCGCTGCG 597  
 Db 538 ATGTGTGCAAAACGCGCTGATCTGCTGCGGCTGACCTGCAATCTGCGCTGCGCTGCG 597  
 Qy 601 TTTATCTGCGGCGACGATTCGCGGCGGAGAGCTGAGGAGCTGAGGAGCAACCTTTATCCG 660  
 Db 598 TTTGCTCTGGGCGCACTGCGGCGGAGAGCTGAGGAGCTGAGGAGCAACCTTTATCCG 657  
 Qy 661 CAGCGCGCGAGAGAGGCTGACAGAGCTTGGGCGCATGATTTGAGGCGCGCTGATGATGAT 720  
 Db 658 CCGGCTGCGGAGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 717  
 Qy 721 GGGCTCTGGGCGAGATTCACCTTATGATCCGCTGAGCAACCTTGTGCTCATGAGCTGAGC 780  
 Db 718 GGGCTCTGGGCGAGATTCACCTTATGATCCGCTTGTGCTCATGAGCTGAGC 777  
 Qy 781 GGGCTGCGGCGCAGCGCTGCGAAACCAAGCGGCGGAGCGCGGCGGAGCGCTGATGATGAT 840  
 Db 778 GCTGTGGCGCGGCTTGGAAACCAAGCGGCGGAGCGCGGCGGAGCGCTGATGATGAT 837  
 Qy 841 AAGACCGCGCTTGGCGGCTGCGGAGAGCGGCGGAGATTTGCTGGGCTGACAGACCGCTG 900  
 Db 838 AAGACCGCGCTTGGCGGCTGCGGAGAGCGGCGGAGATTTGCTGGGCTGACAGACCGCTG 897  
 Qy 901 CCGCGCGAGCACTGGGAGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960  
 Db 898 CCGCGCGAGCACTGGGAGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 957  
 Qy 961 CAACCTCTGGCGAGATGAGGCTGCGCGGCTGATCAACCTCAATCTGGCGAGCGGCGAG 1020  
 Db 958 CAGCGCGAGCTGATGAGAGCGGCGCTGCTGATCAACCTCAATCTGGCGAGCGGCGAG 1017  
 Qy 1021 CCGCGCTGCTGAGCGGCTGCGCTTGGCAAGACCGGAGCATGATGATGATGATGATGATGAT 1080  
 Db 1018 CCGCGCTGCTGAGCGGCTTGGCAAGACCGGAGCATGATGATGATGATGATGATGATGAT 1077  
 Qy 1081 TCGGCGCAATTTCTGTGGGCGCTGATCAACCTCAATGATGATGATGATGATGATGATGAT 1140  
 Db 1078 ACCGGCGAATTTCTGTGGGCGCTGATCAACCTCAATGATGATGATGATGATGATGATGAT 1137  
 Qy 1141 GAGACCGGCTTGTGAGCGGAGAGAGATGCGGCTGTAAGAGCTGAGCGCTTGAATAT 1200  
 Db 1138 GAAAAAGCGCATCTGAGCGCTGAGAGAGTCAAAAGTTTGAACGAGCTGAGACCGCTTAT 1197

Qy 1201 GAGCTGCGCGACCTTCTGGGCTGGGCGGAGCTGCTGCTGAGCGCACTGAAACCGGAGC 1260  
 Db 1198 GAGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1257  
 Qy 1261 ACCGGCATTTACTTTCTTGGCGCTGAAACATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320  
 Db 1258 ACCGGCATTTACTTTCTTGGCGCTGAAACATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1317  
 Qy 1321 GAGTTTACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380  
 Db 1318 GAGTTTACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1377  
 Qy 1381 GAAATATGCGCGCGCATGAGAGGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
 Db 1378 GAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1437  
 Qy 1441 GAGCGCGCTGCGCGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500  
 Db 1438 GAGCGCGAGCTGCGCACTGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1497  
 Qy 1501 GCGCGGACCGACCGCTATTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560  
 Db 1498 GCGCGGACCGACCGCTATTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1557  
 Qy 1561 GCGCGCTTTCGCGAGCTGCGGAGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620  
 Db 1558 ACCCGTGTGCGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1617  
 Qy 1621 TATATGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680  
 Db 1618 TATATGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1677  
 Qy 1681 GAGCGAATGATTTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740  
 Db 1678 GAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1737

RESULT 6  
 AB046580 2913 bp DNA linear ECT 27-JUL-2001  
 LOCUS  
 DEFINITION  
 Pseudoglucanobacter saccharoketogenes adh gene for alcohol  
 dehydrogenase, complete cds.  
 AB046580  
 KEYWORDS  
 AB046580.1 GI:15022438  
 ACCESSION  
 VERSION  
 SOURCE  
 ORGANISM  
 Pseudoglucanobacter saccharoketogenes  
 Bacteria; Pseudoglucanobacter.  
 REFERENCE  
 1  
 Shibata,T. and Saito,Y.  
 Alcohol dehydrogenase  
 Unpublished  
 2 (bases 1 to 2913)  
 JOURNAL  
 TITLE  
 Authors  
 Submitted (26-JUL-2000) Takashi Shibata, Fujisawa Pharmaceutical  
 Co., Ltd., Fermentation Development Laboratories, Shinkawa-cho  
 Nakagawara 156, Nishikawagai-gun, Aichi 452-0915, Japan  
 (E-mail: takashi.shibata@po.fujisawa.co.jp, Tel: 81-52-401-0275,  
 Fax: 81-52-400-1380)  
 FEATURES  
 Location/Qualifiers  
 1..2913  
 /organism="Pseudoglucanobacter saccharoketogenes"  
 /mol\_type="genomic DNA"  
 /strain="IFO 14464"  
 /db\_xref="taxon:133921"  
 209..2035  
 /gene="adh"  
 209..2035  
 /gene="adh"  
 /codon\_start=1  
 /transl\_table=1

source  
 gene  
 CDS



```
/product="alcohol dehydrogenase"  
/protein_id="BAB62258.1"  
/db_xref="GI:15022439"  
/translation="MRFEYLROVNVGLALSTLIALSLGEPAPAOHDANAAAPSRXAG  
SATENROPYADLAKNPANPILRGNYQGWGYSPLDQINQDGLQIOWSRTPMP  
GSNEGAAIANYGTYFLIGNRDVYQALDGTGSLINRYRKLPSAKFTINSLGAARST  
ALFGKDVYFVSWNFVALDARTKGLAMHINQGVESVANSSGTVVDGVYIAGST  
CQSPGRCYVLTGIDASGEELMRNTEIPRGEGDDTVGSEVAMSGQIITD  
PELDLYVSGTSGAPASEVORGETEGTLAGTNRFAVEXKTEYEMRMKQITLPRWMD  
ECTFEMVSTSVNPDAKADGMSVAGNPRGSTRVLGPECKTGVAMOPDAKTDY  
FMKATVEONSLSIDDTGLVYEMDMILKBPKTYNCPFTLGGRDMSAGTLPKSN  
LYIPLSNACVYMARTTEATPADYNTDATTIYLAGKTNMRVADLDAIENETMSY  
ETRAALYDVLTTGGDLVFEVGGIDRDRLIDBSGKEVSTPLPGAVSGYTTYSIDG  
ROYAVAVSGSLGGFTFPPTPDVDSAGANGIYVFALEPK"
```

## ORIGIN

```
Query Match      36.0%; Score 625.8; Db 1; Length 2913;  
Best Local Similarity 61.8%; Pred. No. 9,5e-105;  
Matches 1066; Conservative 0; Mismatches 647; Indels 12; Gaps 4;  
  
QY 24 CACCAAGCGGCGCTGCTATTGCTTAACCGCGCCCGCATTCGCCAGGTAAACCCGAT 83  
Db 307 CAACGCGCGCTGCCGACCGACGACGCGCGCCAGAGCGCTATCGAAGACTTCAGCGCGT 366  
QY 84 TACCGATGAACCTGCGCGAACCGCGCGCTGCTGAATGGAATTAACTACGCGCGCAACCA 143  
Db 367 CACGCGCTGACGATCTCGCGCGCAAGAACCGCGCAACTGCGCGATCTCGCGCGCTACTA 426  
QY 144 AGAAAACTATCGCACTCGCCCTGACCCAGATCACTGCGCAACAGCTTGCTAGTTGCA 203  
Db 427 CAGGCGTGGGGTTACTCGCGCGCTCGACGATCAACAAGAACAGTTGGGCACTTCA 486  
QY 204 ACTGCTCTGGGCGCGCGCGGATGAGAGCGGGGCGCTCAAGTCAAGCCGATGATCATTA 263  
Db 487 GCTGCTGTGTCGCGACGATGAGCGCGGCTCAAGAGGTGCTGCATGCGCTATTA 546  
QY 264 TGGCGTGAATGATCTGGCAAAACCGCGTATGATTCAGCGCGCTGATGCGCAACAG 323  
Db 547 CGCGCTCATCTTCTGCGGAACACCAACAGCTGATCCAGCGGATGAGCGGCAAGCGG 606  
QY 324 CGATCTGATGCGGAACACCGCGCAACTGCGCGCGCTGCGCAACG---CTAAACGCCCA 380  
Db 607 TTCCCTCATCTGGGAATATGTCGCAAGCTGCGCGCTCCAACTTATCAACTCGCT 666  
QY 381 AGCGACCGCGAACGCGCGCTGCGCTTACGCGACGAGCGCTTATTCAGCTCAATGGA 440  
Db 667 CGCGCGCGCAAGCTTGCATGCACTTTCGCGCAAGGTCTATTTGTCGTGTCGGA 726  
QY 441 CAACCATCTGATCGCGCTGATATGAGACGCGCGCAAGTGTATTCGATGTCGAACGTGG 500  
Db 727 CAACCTTGTTGTCGCGCTCGATCCAAAGACCGCGCAAGCTGCGTGGGAACCAATGCGG 786  
QY 501 ATGCGCGCAAGAGGCTTGACAGCTAACCAACGCGCGCGATGTCGCCCAATGCGCTCAT 560  
Db 787 CCGGCGGTTGAAGAAAGCGTTCGAATCTTCGCGCGCATGTGTGATGAGCGCTGT 846  
QY 561 CGTCGCGGTTCCACCTGCGCAATATTCGCGCTATGAGTGTATTCGCGCGCAAGATTTC 620  
Db 847 GATCGGAGGCTCACCTGCGCAATTCGCGGCTTCGCTGATATGTAACCGGATACGACGC 906  
QY 621 CCGGACGCGGTGAGAGAGCTGTGCGCAACCACTTATTCGCGACCGCGCGGAAGAGGTGA 680  
Db 907 TGAATCCGCTGAAGAACTGTGCGCAACCTTCATCCCGCTCGCGGGAAGAGGTGA 966  
QY 681 CGAGACTTGGGCAATG---ATTTCAGGCGCGCTGATGACGCGCGCTGCGGCTGACAT 737  
Db 967 CGACACCTGGGCGGCGCGCGCTTACGAAACCGTTGATGACCGGTCCTGGGCGCAGAT 1026  
QY 738 CACCTATGATCCCGTGACGAACCTTGTGTTATGCTGACCGCGGTGCGGCCACAGCTTC 797  
Db 1027 CACCTATGATCCGGAACCTGACCTGTGTTACTACGCGCTGACCGCGCGCGCGCGCTTTC 1086  
QY 798 CGAAACCGACGCGGACGCGCGGCGGACGCTGTATGCGACCAACACCGCGCTTTCGCT 857
```

```
Db 1087 GGAAGTCGACGCGCGACCGAAGCGCGCACTCTGACAGGACCAATACCCGCTTTCGCT 1146  
QY 858 GCGTCCCGACACGCGGCGGATTTCTGCGCGCTCACCAAGCCCTGCGCGCGCAACCTGGGA 917  
Db 1147 GAGCGCCAGACCGGTGAAGTCCTGGAAGACCAACCTGCGCGCGCAACCTGGGA 1206  
QY 918 CCAAGATGACGCTTCGATGATGTCGCCAAGTGCATGTCGAACCTTCGCGCGAGAT 977  
Db 1207 CTCGGAATGACGCTTCGAATGATGTCGTCGACCTCGTCAACCCGCGCAAGCG 1266  
QY 978 GGAAGGCTTCGCGCGCCATCAACCCCAATGCGCGCGAAGCGGGA---GGCGCGTGCAGC 1034  
Db 1267 CGATGCGATGATGTCGTCGTCGCAACGTCGCGCGCGCAAAACCGCGCAAGTGCAGC 1326  
QY 1035 GGGTCGCGCTTGCAGAACCGCGACGATGTCGCTTGTGATGCGCGCTTCGCGCAATTCCT 1094  
Db 1327 CCGGCTGCGCTGCAGAACCGCGCTGCGCTGCACTTCGATGCGCAAGACCGCGCACTACTT 1386  
QY 1095 GTGCGCGCGTATACCACTACCAATATGATGTCCTTCATGCAAGACCGCGCTTGT 1154  
Db 1387 CTGCTCAAGGCAACCGTCGACAGAACTGATGCTTCATGATGACACAGGCGCTGT 1446  
QY 1155 GACGCTGAACGAGATGCGGTCGTAAGAGCTGAGCTGATGATGATGATGATGATGATGATGAT 1214  
Db 1447 TACGCTCAATGAGACATGATCTCAAGAGACCGGCAAGACCTAATTAATGCGCGAC 1506  
QY 1215 CTTCCTGCGTGGGCGCACTGTCGTGTCAGCGCGCACTGAACCGCGCAACCGCATTTACTT 1274  
Db 1507 CTTCCTGCGCGTGTGTCGTCGTCGCGCGCGCTGACCTGCGCAAGTCAACCTCTACGT 1566  
QY 1275 CTTCGCGCTGAACATGCTGCTGATGATTAATGAGCGCTGATCAAGAGTTTACGCGCT 1334  
Db 1567 GATCCCGCTCAGCAACCGCTCTCTACAGCTGATGATGATGATGATGATGATGATGATGAT 1626  
QY 1335 CGACGCTATTAACCAACGCGGACCGCAAACTCGCGCGCGGCTTTGAAATATATGAGCGG 1394  
Db 1627 TGACGCTTAACCAACCGGACCGCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1686  
QY 1395 CATGACCGCATTAATATCAACACCGGCGGCACTTGTGTCGCGCGGACCGCGCTGCGCG 1454  
Db 1687 CGTGAATGCAATGATGTCGCAACCGGAGAGCAAGTGTGCTTACGAACACCGTGTGCGG 1746  
QY 1455 GAACCTACTCGCGCTTTTGTGACGAGGAGCGGCTGTGTTCAACGCGCGGACCGACG 1514  
Db 1747 TCTCTAGACCGCGGCTCTGACGACCGCGGAGCACTGTGTGTGTCGCGGTAATGATG 1806  
QY 1515 CTATTTCCGTGCTCCTCAGCGAGAAACCGCGGAGACTTTGTGCGAGGCGGCTCTTGCAGC 1574  
Db 1807 TGACTTCCGCGCTCTGACGCGCGGATCCGCGCAAGAGTCTGTGTCACCCGCTCGCGG 1866  
QY 1575 GGTGCGACGCGGAGGAGGATCAGCTACAGTTGAGCGCGCTGCAATATATGCGCATCGG 1634  
Db 1867 TCGGCTTCCGCTACACACCACTACTCATGATGATGATGATGATGATGATGATGATGATGAT 1926  
QY 1635 TCGCGCGCTGACCTATGAGCAAGCAATGAAACGCGCGCTGCGCGCAAGCAATGATTC 1694  
Db 1927 CTCGCGCGGCTCGCTCGCTGCGCGCGGACCTT---CGGCGCGCAACCCCGCGCGCTGCTC 1983  
QY 1695 GACCTCGGTCGCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1739  
Db 1984 GACTTCGCGCGCAACCGCATCTACGTCCTGCTTTCGCGAGAA 2028  
  
RESULT 7  
AP005957  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
Bradyrhizobium japonicum USDA 110
```

ORGANISM	Bradyrhizobium japonicum USDA 110
REFERENCE	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Bradyrhizobiaceae; Bradyrhizobiaceae.
AUTHORS	1 Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Matsumoto, M., Iriuchida, M., Kawashima, K., Kohara, M., Matsumoto, M., Shimo, S., Tsuruoka, H., Wada, T., Yamada, M. and Tabata, S.
TITLE	Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110
JOURNAL	DNA Res. 9 (6), 189-197 (2002)
PubMed	22484998
REFERENCE	12597275
AUTHORS	2 Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Matsumoto, M., Iriuchida, M., Kawashima, K., Kohara, M., Matsumoto, M., Shimo, S., Tsuruoka, H., Wada, T., Yamada, M. and Tabata, S.
TITLE	Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110 (supplement)
JOURNAL	DNA Res. 9 (6), 225-236 (2002)
PubMed	22485002
REFERENCE	12597279
AUTHORS	3 (bases 1 to 299910)
TITLE	Direct Submission
JOURNAL	Submitted (20-NOV-2002) Takakazu Kaneko, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research; 2-6-7 Kazusa-Kametari, Kisarazu, Chiba 292-0812, Japan
REFERENCE	(E-mail: kaneko@kazusa.or.jp, URL: http://www.kazusa.or.jp/rhizobase/, Tel: 81-438-52-3938 (ex. 2338) Fax: 81-438-52-3934)
FEATURES	Location/Qualifiers
source	1..299910
gene	/organism="Bradyrhizobium japonicum USDA 110"
CDS	/mol_type="genomic DNA"
gene	/strain="USDA110"
CDS	/db_xref="taxon:224911"
gene	7..1380
CDS	/gene="blr6001"
gene	/note="ORF ID:blr6001"
CDS	unknown protein"
gene	/codon_start=1
CDS	/transl_table=11
gene	/protein_id="BACS1266.1"
CDS	/db_xref="GI:27354279"
gene	/translation="VVDHFKRDLPGAIVMLGYLARQSVVALVPMYEQADVPRGLD
CDS	ALVNYARPVNLDLNRSPFAKAGIALVLDTEGQVLAEGKNSPEMAAAVKGSGYADI
gene	LGVYFQGNRLHDAFLTAGTMRBQHLTGCRFPDPAARWMLADGEPQGLVNNAN
CDS	PLVNGRFTGKPGGEREAMVAVGADGVYDRIADLKQVFAVLAIDRLAAAPDRIT
gene	ILVRPPESEDDYRNALAHANVVIDGTGSVLDRIKNAAVVHLNCGAAVESVILGK
CDS	LPLOLEYNTPTTGAHALPARVSSVNSFDLLGAIIDHIERETETPDFAVHADIE
gene	APFHLDGQAAERVADVLTDAANSRRPYSILATYKGTDRKSIQIYKGAASAVLGS
CDS	AVTERLSRFPNRKRIEVPVFTLLORIAHDPASPAQFTATRAHCATTGLPLAS
gene	IALRPAR"
CDS	1384..2787
gene	/gene="blr6002"
CDS	1384..2787
gene	/note="ORF ID:blr6002"
CDS	unknown protein"
gene	/codon_start=1
CDS	/transl_table=11
gene	/protein_id="BACS1267.1"
CDS	/db_xref="GI:27354280"
gene	/translation="MATSARITLITLTAAYOTRFMTVPAQRLRAAGHVELLAPDRS
CDS	AEKSAEGVPVTMYREGKAGSPSPDRGAFPARVAAGLDTNLFSEHETPTGIRD
gene	TLALRRFMTYANAEVTLDRJBGHGQALVQEGGSPVLAFTYARRRITRMFI
CDS	EBSPFRGRMYFTPDITGADVMSPADVSPVRAVLDITLIRAIIVPKDQHYSA
gene	AKKKVNLNRNARLTTEKWDQALGQGFGRHLRHARVHAAMALNARLRLTYPIP
CDS	DAFVYVPHVADMALTLRSPDYLDQVATVDFVLTITPDSHVLVKEHPAQIGAIISA
gene	DRULELARRPDRNFVLLPQTNNTVTLNRADAVTSVNSKGAALILGKVVYNGDAFY
CDS	RSCPLVYADVRLADVPAURLREALSAGPDPARQAPYFESAMSSSYGELIYIDPKLLD
gene	TFAASLRRAAIAERPARVN"
CDS	2799..3548
gene	/gene="blr6003"
CDS	2799..3548
gene	/note="ORF ID:blr6003"
CDS	probable glycosyl transferase"
gene	/codon_start=1
CDS	/transl_table=11
gene	/protein_id="BACS1268.1"
CDS	/db_xref="GI:27354281"
gene	/translation="MPLVSLITPSNVEGLIERTISNOSQTFADWELLIDDCSDR
CDS	TPETIADIGEDRVRKILRMPKNGPALRQAHIDRQRTAFLDSDLMFGKER
gene	QLMPARKRALSLYTLARRINENTVGRILISVASTYGLDKNTAINTLVADRS
CDS	IACPIVKNKGVDYDFCLMLSLIKRGHTAGLDLARYVYRGSSVSSRVPVRAKWWQ
gene	YRNVEHLISLVRSSMCFPAHWGARAMLKREEF"
CDS	3706..4668
gene	/gene="blr6004"
CDS	3706..4668
gene	/note="ORF ID:blr6004"
CDS	/codon_start=1
gene	/transl_table=11
CDS	/product="dUDP-glucose 4-6-dehydratase"
gene	/protein_id="BACS1269.1"
CDS	/db_xref="GI:27354282"
gene	/translation="MPESTKSRILIVTGAGFGISGICERLLDAGAVSADNPTG
CDS	SRNNIHLIANPLFEAVRDVTFPLIYEDALFNLCPSPIHQSDPVQTKTISVHG
gene	AINMLGLAKLILKARIFQASTSEVYGDPLHQEDYDGNVNPIGISCYDEBKRCET
CDS	LFEDYVRQGLPIKVARIFNTYQPRMGPDGRVSPFIYQALQGEPIITVGGGQGRS
gene	FCYVDIIVERIKRLMTKEDITGPINIGNSSEITIELAEKVIETLTSKLVFKXLP
CDS	ODDPRORQDPLTKAKTALMPEKVALMEDGKETIAYFKSLSLA"
gene	4711..6024
CDS	/gene="blr6005"
gene	4711..6024
CDS	/note="ORF ID:blr6005"
gene	unknown protein"
CDS	/codon_start=1
gene	/transl_table=11
CDS	/protein_id="BACS1270.1"
gene	/db_xref="GI:27354283"
CDS	/translation="MSAPSGIMQHLQSRDLSKRALITLLLSAVTCLSLIKVATFOS
gene	YIHSGERWIAVAVAVAFSVSLFPAARFSFGAGFYFMILGFMIDVFSYS
CDS	YDRLLAGLSAASLVLLLEPLVRAPFQQLVLSNAGEHLITALLVSVGTIAVAS
gene	TNPRILVAIANIYDARDALAPGAVYILIGVSSSTLLPTTFACYSILGMRAGVLI
CDS	ILLFPVYITTKRAFPFPKMLITLAVLSSELTERTSVITSIFFYPMILGLVIAVGA
gene	INSPLGKYPDLVITRMATTSALDITNHFPAHPIITWCQISMLKPNVPCAYODLS
CDS	VWENNTYFQMIINASLPATEGVASVGPYALPLTALVAGFTLAGNPSAGLPQRLVLI
gene	SSQTLPHVLTNLEPLTVAMVHGTAFLFLIMYVPSIFQPGLPD"
CDS	complement (6021..7199)
gene	/gene="blr6006"
CDS	complement (6021..7199)
gene	/note="ORF ID:blr6006"
CDS	probable glycosyl transferase"
gene	/codon_start=1
CDS	/transl_table=11
gene	/protein_id="BACS1271.1"
CDS	/db_xref="GI:27354284"
gene	/translation="MPEAPRIATFTNIPRIIMAGVYVORNLFAIHRHSPGTAVV
CDS	LFAGTRDDPEELALASITRAVALVRSPADDRKTCGLGALACRGPALAPPAQVD
gene	ITVESARFQGMRLPCPALAMPDQRLILPLFSRAVYRRDQGRQAQASGHNLIS
CDS	SASALGDLKRFPGISNGVSVFATSPVSLATKPSVIAQYGLPSYFYINPFM
gene	RHKHNGVVDALGILKGRGIDVVAAGSTRDREDFEPTMSEVRSGLTENFRHL
CDS	GKMLPDLHVALLASMALINPSECEGMSITVEAKSGFVPMIISDIDVHREQTEGAR
gene	YFGIRDAELAAHLSSVAGSADPTIRDLPLDEERVKAFVADPFRVLTADVQNSGIR
CDS	complement (7192..8124)
gene	/gene="galE"
CDS	complement (7192..8124)

```
/gene="gale"
/ncore="ORF_ID:b116007
probable UDP-glucose 4-epimerase (EC 5.1.3.2)"
/codon_start=1
/transl_table=11
/protein_id="BAC51272.1"
/db_xref="gi:27354285"
/translation="MSAWTGANFGRHLVRELAVGRTVHGSHGALDPAEARALG
LOTINGEVAANLSALATHG.PSOIFHLAGSSVGLSTARPEDEPSFTSTARLT
EMRSPARESLIVASAAVYGDHAPISASALAPMPSYGHKMKMQLCRSYAQS
FGHCTVRLFSYVYENRKLMDICSRUSSESLNIGTAEIRDTVDVRL
LVCAAEVQDEDFRLINGSGRGTSVAELIRWGSSTVYRSGVAPDGPASLT
ADGRLDMNFDRLLPLERGLADYEMFRSQARA"
complement (8205 .8825)
/gene="b116008"
complement (8205 .8825)
/gene="b116008"
/ncore="ORF_ID:b116008
putative acetyltransferase"
/codon_start=1
/transl_table=11
/protein_id="BAC51273.1"
/db_xref="gi:27354286"
/translation="MSLSLTIVRCVHAIDENFLATVQLIGRATCRIDRGAFLAH
ARLNALGDEKIVIGSHSHVRELMTLHGQINIGECYGVGRISWASIDIGN
RVLISHSVNIFDLNLSHPIRASERHEQAKELFTGHPROLSDRPRIKICDDMIGACA
MVRGVYIGEGGIVAAVAVTKDVPAYSVIAGNPATIVRELSPDER"
complement (8903 .9991)
/gene="b116009"
complement (8903 .9991)
/ncore="ORF_ID:b116009
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAC51274.1"
/db_xref="gi:27354287"
/translation="MRIFQNSCFYPSYLPRLNOLAANKAVSFERRRVLNDRFASHF
LQPLVDGSPFAFTNGDDDEVLRMARBEQMGEPFLLEILAQLEHAEVFLYLDP
VRYPSAFVALPGCVKALCMRGAPEGNADFTAYGVLANFPMLEMTMRKCRLELF
VPADHVMDEYGERPIDVYFVGSGSRHREKATLEVARLADSHVFLCIDSRL
TKLAESALIGRLPLRQHRPDAIAIARPEVPRDIYELIGSAKILNDSIDMAGDR
GNMRCPESMGCALLVSDAGRTDMESGITITETYPDPAHAEVLARSIDMPRSEI
AALGPARIIDYHKMDQKLFYVVGRL"
complement (9995 .11835)
/gene="b116010"
complement (9995 .11835)
CDS
Query Match 8.1%; Score 141.8; DB 1; Length 29910;
Best Local Similarity 47.3%; Pred. No.2.9e-16;
Matches 714; Conservative 0; Mismatches 737; Indels 60; Gaps 7;
```

```
Db 225140 TCTTCCGACACACCTTCAGACGCGCAACGTGATGCGCTCGACCCAGGACGGCAAGAGC 225199
QY 482 TATTCGATGTGCAACGTGATCGGGGCGAAGCGCTTGACCAATGACACACGCGGCGCA 541
Db 225200 TTTGGCGCGCAGAAAGCGCGCCATATCAAGAAAGGCTATTTCATG--ACGGTGGCGCCGC 225256
QY 542 TTGTGGCAATGAGGCTCATGCGCGGGGTTCCACCTGCAATATTCGCGCCATGATGATGCT 601
Db 225257 TGGTCGCGGACGCGCTGCTCATACCGGCATCTCCGGCGCGGACGTTGGGACCCGCGCT 225316
QY 602 TTATCTCGGGGACGATTCGCGACGCGTGAAGACTGTGCGCAACCACTTATCCGC 661
Db 225317 TCATCGACGCGTGGGATCCGCGACGCGCAAGCATCTCTGCGGACCCATTCGATCCCT 225376
QY 662 AGCCGCGGAAAGAGGTACAGAGACTTGGGCGCATGATTTGAGGCGCGCTGATGACCG 721
Db 225377 CTCCGACGAGCCCGGTGCGACACCTGGAAGGCGACACTGGAAGCTCGGCGGCGCT 225436
QY 722 GCGTCTGGGTCAGATCACTATGATCCCGTGAAGCACTTGTGTTCTATGCTCGACCG 781
Db 225437 CGACTTGATCAACGCGCTCTACATCCGATCCGAGCTGGAACAGGTCTATTGGGGCATCGGCA 225496
QY 782 GCGTGGGCCACAGCTCGCAAAACCAAGCGGCGACGCGCGGCGGACGCTGATGGCACCA 841
Db 225497 ATCCCGGCGCTTCACACTCGCGGCGTGC-----CCGCGGACCACTCTA---CACT 225547
QY 842 ACACCGCTTTGCGGTGCGTCCCGACACGCGGAGATGTCTGGGCGTACCAGACCCCTGC 901
Db 225548 GCTCCGTGCTGCGATGATCCCAAGACCGGCAAGATCAATGGCACTACAGTTCTTGC 225607
QY 902 CGCGGACAACTGGGACCAAGATGACAGTTCAGATGATGTCGCCAACGTGATGTC 961
Db 225608 CGAACAATCCGTTGCACTATGACAGCGGCGCGAGATGCTCGCGCAATGAACGTCG 225667
QY 962 AACCTCGGCGAGATGGAAGGTCGTGCGCGGCAATCAACCCCAATGCGGCGGCGAGC 1021
Db 225668 AGGGAACCCGACCAAGGTGCTGATGATGCCAAC----- 225703
QY 1022 GCCGTGCTGACGCGGTGCGCTTGCAAGACCGGACAGATGTCGTTGATGCGGCT 1081
Db 225704 -----GTAAAGGCTTCTTCTAGTGTCTGACC-GAACCAAGAAAGTGTCTCGGGCCA 225757
QY 1082 CGGGGAAATTCCTGTGGCGCGGTATACCACTACCAATATGATGCGCTCGATTCAGC 1141
Db 225758 ATCCCTACGTAAAGTGAACGTGGGACCGGCGTCGACATGAACAGGCGCGCGCATCG 225817
QY 1142 AGACCGGCTTGTGACGCGTGAACGAGATGCGGTGTAAGAGCTGGAAGCTGGAATATG 1201
Db 225818 AGACCGAGCTTTCAAGAGACCGCGGAGGCAAGAGTACGTCTATTCGCTGATCC 225877
QY 1202 AGCTGTGCCGACCTTCTGTGGTGGCGCGCATGCTGTGTAGCGGCACGTGAACCCGACA 1261
Db 225878 TCGGCGGCAAAACGTGGAGCGCATGTCTCAATCCGCAAAACCGGATTCGCTTACGCA 225937
QY 1262 CCGGCAATTAATTCCTGTGCGGTGAACATGCTGTCTACGATTAATGCGCTGATCAAG 1321
Db 225938 ACAGCTGCGCTTTCGCGCGCAATGACAGGCGCAACCGGTACCTTCAAGAGGCGAAT 225997
QY 1322 AGTTAGCGCGCTGACGCTTATTAACACAGCGCGACCGCAAACTGCGCGCGGCTTTG 1381
Db 225998 GGTATCTGCGATGACCTTACCGATCCCTGGAAATTCGSGGACGCGCGGCGGCGCATC 226057
QY 1382 AAAATATGGCGCGCATGACGCGGATTAACATACACACCGCGCGGACCTTGTGTGCGCGG 1441
Db 226058 TGAAG-----CGATCGATCCGATGACCGGCAAGCGCAAGTGGAGGCGC 226102
QY 1442 AGCGCCCTGCGGCAATCACTGCGCGCTTTGTGCGAGCGGCGGTGTGTTCAACG 1501
Db 226103 CGCGGACATCCGCGGCTTCTCCGCGGTGTGTGACCGCGGCGGCGGTGTGTTCAACG 226162
QY 1502 GCGGAGCGGACCGCTATTTCCGTGCTCGACCGAGAAACCGGCGGACCTTGTGCGAGG 1561
```

Db 226163 GCGCCCTGACGGCGAGTTCCGATCCGACACCGGCAAGACCTCTGGCACT 226222  
 1562 CCCGTTTGGACGCGTCCGACCGGCGAGCGATCAGCTTACGATTTGACGGCGTGCAT 1621  
 Db 226223 TCCAGACCGGCTCCGCGATCGAGGCAACCGGTGACGTGACGAGATGACGTCAAT 226282  
 1622 ATATGCCCATC 1632  
 Db 226283 ACAATGCGCTC 226293

RESULT 8  
 AP005957/c 299910 bp DNA linear BCT 12-SEP-2003  
 LOCUS Bradyrhizobium japonicum USDA 110 DNA, complete genome, section  
 DEFINITION 23/31..

ACCESSION AP005957  
 VERSION AP005957.1  
 SOURCE GI:27354278  
 KEYWORDS Bradyrhizobium japonicum USDA 110  
 ORGANISM Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 Bradyrhizobiaceae; Bradyrhizobium.

REFERENCE 1  
 AUTHORS Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiyama, T.,  
 Sasamoto, S., Watanabe, A., Idesawa, K., Iriyochi, M., Kawashima, K.,  
 Kohara, M., Matsumoto, M., Shimpou, S., Tsuruoka, H., Wada, T., Yamada, M.  
 and Tabata, S.  
 TITLE Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
 Bradyrhizobium japonicum USDA110  
 JOURNAL DNA Res. 9 (6), 189-197 (2002)  
 MEDLINE 22484998  
 PUBMED 12597275  
 REFERENCE 2  
 AUTHORS Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiyama, T.,  
 Sasamoto, S., Watanabe, A., Idesawa, K., Iriyochi, M., Kawashima, K.,  
 Kohara, M., Matsumoto, M., Shimpou, S., Tsuruoka, H., Wada, T., Yamada, M.  
 and Tabata, S.  
 TITLE Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
 Bradyrhizobium japonicum USDA110 (supplement)  
 JOURNAL DNA Res. 9 (6), 225-256 (2002)  
 MEDLINE 22485002  
 PUBMED 12597279  
 REFERENCE 3  
 AUTHORS Kaneko, T.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-NOV-2002) Takakazu Kaneko, Kazusa DNA Research  
 Institute, The First Laboratory for Plant Gene Research; 2-6-7  
 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
 (E-mail: kaneko@kazusa.or.jp / rhizobase/  
 URL: http://www.kazusa.or.jp/rhizobase/  
 Tel: 81-438-52-3935 (ex. 2338), Fax: 81-438-52-3934)

FEATURES  
 source  
 1..299910  
 /organism="Bradyrhizobium japonicum USDA 110"  
 /mol\_type="genomic DNA"  
 /strain="USDA110"  
 /db\_xref="taxon:224911"  
 7..1380  
 /gene="blr6001"  
 7..1380  
 /gene="blr6001"  
 /note="ORF ID:blr6001  
 unknown protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BACS1266.1"  
 /db\_xref="GI:27354279"  
 /translation="MVDHPRDLPAGVMLGQYLARGVSVLVPMYEQAVDVPRLGLD  
 ALVVNARPNLIDMSPAKAGLALVLDTEGVLAEGKNSPPMAAHLVKSQYADI  
 LGVFFNGRILHDAFLTAGMKREQLHGTGCRPDPAARMRALDGEPRGLVNNAN  
 PLVNSRFGKCGGREGAEMVRGMDAAYVDRIADI KOYFANYLAEIDRIAAARPDRT  
 ILVRPFESSEEDYRNALAHANVYIDGTGSLDRIIRNAAYVHLNCGIAVESVLGK

gene  
 CDS  
 7..1380  
 /gene="blr6001"  
 /note="ORF ID:blr6001  
 unknown protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BACS1266.1"  
 /db\_xref="GI:27354279"  
 /translation="MVDHPRDLPAGVMLGQYLARGVSVLVPMYEQAVDVPRLGLD  
 ALVVNARPNLIDMSPAKAGLALVLDTEGVLAEGKNSPPMAAHLVKSQYADI  
 LGVFFNGRILHDAFLTAGMKREQLHGTGCRPDPAARMRALDGEPRGLVNNAN  
 PLVNSRFGKCGGREGAEMVRGMDAAYVDRIADI KOYFANYLAEIDRIAAARPDRT  
 ILVRPFESSEEDYRNALAHANVYIDGTGSLDRIIRNAAYVHLNCGIAVESVLGK

gene  
 CDS

gene  
 CDS

gene  
 CDS

gene  
 CDS

gene

LPLOLEYLNTPTAGHALPARSVRSVNSFDELLGAIIDHIERETETFEAPVAAADIE  
 APFHINDGOAERVADVLDIDANSRRPYSLATVGTGDKDSIGQIVGAASAVGS  
 AATERRLSRNPARRDKRIEPIYFVRLILRLIAHDIASPAQFTATRAHCAITGLPLAS  
 IALBPAA"  
 1384..2787  
 /gene="blr6002"  
 1384..2787  
 /gene="blr6002"  
 /note="ORF ID:blr6002  
 unknown protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BACS1267.1"  
 /db\_xref="GI:27354280"  
 /translation="MATSARTILITLAEYQTFWIPVAQRRLRAAGHDELLAFDDRS  
 AEMSAAGVPTVMYREGKAGSPDDRGAFDARVAAGIDGNPLFSHERPTGIARD  
 TALRRPMTVNTAMMETVLDLEGHKQAEIVOEIGPISLVASPYAARRNINMFI  
 EPSFRGKRTFTDITGADPVMSPADVYSPEVRALIDPTLQRAVIRKQOHVSA  
 AFKKVNLKVARLLEKMDQFALGQESGHNLRRARVHAALNALNLRKLYKPIF  
 DAPFVYVFFVADMALTLRSDYLDQVATVDFLRTIPDSHYLVKEHPAQIGALISA  
 DRUFELARRDNFVLLPQTNNTYVLRADAVISVNSKGAELILGKRVVWGDNFY  
 RSCPLVYAVDRADVAPRLREALISAGFPDARGAPYFESAMWSYSGELVISPDKILD  
 TPAASLRRAIAEAPRVN"  
 2799..3548  
 /gene="blr6003"  
 2799..3548  
 /gene="blr6003"  
 /note="ORF ID:blr6003  
 probable glycosyl transferase"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BACS1268.1"  
 /db\_xref="GI:27354281"  
 /translation="MPLVSTITPSNVEGLIEETISVOSQTFADMBELLIADDCSDR  
 TPEIVADIGERDPRVRLIRPKNGGPAALQAAIDBAQGRYLAFLDSDILWPKLER  
 QLAFAEKRAALSYTAFRIINETNTVGLQISVAPALTYGOLIKNTAIIALVAVRE  
 IAGPIVKNKGVDYDFCWLMSILKRGHTAYGLDRLARYRVRGSSVSRPVSAAKMWQ  
 IYNNVETSLVRSWCPFAHMGARAWLKRRBF"  
 3706..4668  
 /gene="blr6004"  
 3706..4668  
 /gene="blr6004"  
 /note="ORF ID:blr6004"  
 /codon\_start=1  
 /transl\_table=11  
 /product="rmp-glucose 4-6-dehydratase"  
 /protein\_id="BACS1269.1"  
 /db\_xref="GI:27354282"  
 /translation="MPFESYKSRILVITGAGFISGHICERLLDAGAEVVSADNYTG  
 SRRIATLAMPLEFAVRHDVTPFLIEVDALFNLAQCPASPIHYQSDPVQTTKTSVHG  
 AINMLGLAKRLPIFOASTSEYVGPLIHPOTEDEYGMVNPPIGISCYDEGRCAET  
 LFPDYMRHGLPKIKVARIENITVGRMOPNDGRVSSFIYVALGEBITVFGDGQOTRS  
 FCVYDDIYKAIMLMTTKEDIKGPINIGNSFETIELEAKETELGSRSKLVFKLP  
 QDDPRQRPDLTKAKTALNWEPRVALDGLKETITAFKSLSLA"  
 4711..6024  
 /gene="blr6005"  
 4711..6024  
 /gene="blr6005"  
 /note="ORF ID:blr6005  
 unknown protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BACS1270.1"  
 /db\_xref="GI:27354283"  
 /translation="MSAPSGIMQHLQSRLLDSKPALTILLHSAYVCLSLIKVATFOS  
 YIHRSGRWMIAVAVVAPSGVSLFLPAARFSFGYFAGFYETMILGFLMIDVFSYS  
 YDRILACLASAASIVILLIPYFVRAPFQIVYLSSNAGEHLLTALIVSGTIAVAS  
 TYNRLVAINITIDYRDALFPCAAYILIGVSTILPTPRACYSILGRHWRAGVAILI  
 LILFPIYITLTKFAFTPALVITLAVLSRFLKRTSVTISIEVPLILGVITAVTGA  
 IVNSIPGCFPLVILRMATSSALDIYNHFAHPLVLRFCQISMLKLPWPCAYODQUS  
 VMENTYGFEMINASIFATEGVASVGYLAFLTALVAGFILAAGNWSAGLPQRLVILI  
 SSGTLPLVILNTPLTVMVTHGTALLFLIMYVMPRIIFQPGHDP"  
 complement (6021..7199)

```

CDS
    /gene="b116006"
    /complement(6021..7199)
    /gene="b116006"
    /note="ORF ID:b116006
    probable glycosyl transferase"
    /codon_start=1
    /transl_table=11
    /protein_id="BACS1271.1"
    /db_xref="GI:27354284"
    /translation="MPERAPLRAPFNI PRLMAGSYNYQNL FALHRYSPQVAVP
    LEAGTDDPEELALALIPALVLRSPAPDRSATGLGALGRDGPALAEPAQVD
    IVESARPFQMLPCPAIAMPPDPORHLPHLFS88AYRDLGPAPQIASGRHMLS
    SASLGDIKRYPGSLNGSVYFRAPEPVSLIATSPVIAOXGIPSPYPLPMQFW
    RKNHOVVDALGLKRGIDVYVNASGSTRDPRDFERLTERSEYRSGLSTNRHL
    GMLPDHVYALLKASMALINPSCGEMSTVEAKSFVPMIUSDIDVHREQTESAR
    YFGIRDAEALNAHLSEVAQADPFTIRDLLPLEERVKAFVADFVRLTADVWNSGIR
    "
gene
    complement(7192..8124)
    /gene="gale"
    /complement(7192..8124)
    /gene="gale"
    /note="ORF ID:b116007
    probable UDP-glucose 4-epimerase (EC 5.1.3.2)"
    /codon_start=1
    /transl_table=11
    /protein_id="BACS1272.1"
    /db_xref="GI:27354285"
    /translation="MSAWTTGANGTIGRLVRELVAGVTVHGVGHGALDPAEAAALG
    LQWINGEVDAANLSALAAHGLPSQIFILAGSSVGLSIARFEDFSTVSTALL
    EWLRSFAPBSRLIVASSAAYGADHAGPIAESALAPMPYGHKXIMEQLRSYVQS
    FGLHCTVRLFSYVGNLRKQLMDICSLSHERSLNGTGAEIRDTVDVRL
    LVCAAEVQOEDFRLINGSGRGTSVAELAGLIRWGSSTVYRSGVARPDPSALL
    ADGRLBDMFMDRLPLERGLADYVEVFSQARA"
gene
    complement(8205..8825)
    /gene="b116008"
    /complement(8205..8825)
    /gene="b116008"
    /note="ORF ID:b116008
    putative acetyltransferase"
    /codon_start=1
    /transl_table=11
    /protein_id="BACS1273.1"
    /db_xref="GI:27354286"
    /translation="MSLSLTVAMCVHAIDGFMHLYVORLIGRATCRIDRGAFIANS
    ARRNALGDESEKIVIGSHSVRGLMILGHGQINIGWCYGVGTRIGSGSIDIGN
    RVLISHVINFDNLSHPIRASEHEQAKIFSTGHPRDLSIDRPKIQDDAMTGACA
    MVRGVTTIGCGTVAAGAVYTKDVPASIVAGNPATIVAEISPDER"
gene
    complement(8903..9991)
    /gene="b116009"
    /complement(8903..9991)
    /gene="b116009"
    /note="ORF ID:b116009
    unknown protein"
    /codon_start=1
    /transl_table=11
    /protein_id="BACS1274.1"
    /db_xref="GI:27354287"
    /translation="MRLFONSCFPYSYLPRLNQLAAKAVSPERRRVFLINDFGASHF
    LQVLDGSPAPFTNGDDEVLOSRMARBOGMBEPLILALQLEHNAHTEVYNDP
    VRYPSAFVAKLPGCKVKALCMRGASGMDFTATYGVLANPMTLDMWRKRAELF
    VYAPDVMDEYGHGERPIDVVPYGVYSRGNHREAKTLERVAALADSHVYVRCDSRL
    TKAESNIGRLPLRHRPRDAIJAIRPVRERDLYELIGAKIALNGSISIMAGEDR
    GKNRCFESWCGCALVSDAGRYIDGESGTTLETYDTPHAAVLARIQLDMPRAEI
    ALGRRARIRDIYHKDQMKLFEVVGRL"
gene
    complement(9995..11833)
    /gene="b116010"
    /complement(9995..11833)

```

```

Query Match      6.8%; Score 118.2; DB 1; Length 299910;
Best Local Similarity 49.4%; Pred. No. 6.1e-12;
Matches 433; Conservative 0; Mismatches 423; Indels 21; Gaps 4;
QY 50 CCGGCGCCGCGGATTCGCGAGGTAAACCCGATACCGATGAAGTGTGCGGAACCCGC 109

```

```

Db 240677 CCGGTGCGGCCCATCGAAGATTATTCCTCGGTACCGCGCACGCTTGAGATCCG 240618
QY 110 CCGGTGTGATGATTACTACGCGCGCAACCAAGAAACTATCGCACTCGCCCTGA 169
Db 240617 AACGAGCACTGATGCTGCTATGCGGCACCTTATGACGGGCAAGGCTACAGCCGCTCG 240558
QY 170 CCCGATCATCTGCGGACCAAGCTTGATGCAATGTTGCACTGGTCTGGGCCCGCGGATGAGAG 229
Db 240557 ACCAGATCAACACTCGAAGCTGAAAGGCTCAACGCGGTCTGACGTTTGCAACAGCG 240498
QY 230 CCGGGGCGGTAC---AGGTCAAGCGGATGATCCATGATGAGGTGATGATGCGCAACC 286
Db 240497 TCGTCAAGGCAAGAGCGCGCGCGATCGTCAACATGCGTATGTTGCGGCGACC 240438
QY 287 CCGGTGATGATCCAGCGCTGTGATGCGCAACAGCGCATCTGATCTGGACACGCGC 346
Db 240437 CGATGGGCGAGGTGATGCGCGCTGAACGGAAACCGGCGACGATCTGGGCGTACAGC 240378
QY 347 GCCAATGCGCGCGCGTGCACCGCTAAACGCCCAAGCGACCCGGAAGCGCGCTCGCC 406
Db 240377 GCGAGCTCCCGACG-----ATCTGTTCACTGATTCGACCAAGCGCGCGCTCGACT 240324
QY 407 TTACGCGACAGCCTCTATTTCAGCTCATGGAACAACATCTGATCGCGTGTGATATG 466
Db 240323 TGTGGAGAGCAAGCTATATCTCGGCACACCGACGATCANTGTGTCGCGCTGACGCGCA 240264
QY 467 AGACGGGCGCAGTGTGATTCATGATCCAGCTGATCGAGCGGCGAAGAGGCTTGACAGTA 526
Db 240263 AGACGGGCGAAGTGTGTGGACACCAAGGTGACAGACTAACAAAGGCGCGATATGA 240204
QY 527 ACACACGCGGCGCATTTGCGCAATGCGCTCATCGTCGCGGTTCCACTGCCAATAT 586
Db 240203 CCTGATCCGCTGATCTGTCACGCGCAAGTCACTGTGCGGCGCTCCGCGCG---GAGT 240147
QY 587 CGCCCTATGATGCTTATCTCGG3GACGATTCGCGACGCGGTGAGAGCTGTGGCGCA 646
Db 240146 TCGGCGTCCGCGCTATGTCGCCCTCTATGACGCGCAAGAGCGAGCGTGTGGCGCA 240087
QY 647 ACCACTTATCCCGCAGCGCGCGGCGAAGAGGTGACGAGACTTGGGGCAATGATTTCAGG 706
Db 240086 CTTACACCATTCCTCGGCGAAGGCGAGCCCGTTCACGACCTTGCAAGGCGATCACTGGA 240027
QY 707 CGCGGTGANTACCGCGCTCTGG3GATAGATCACTATGATCCGTCGACGAACCTTGT 766
Db 240026 AGAAGCGGCGGCGCTCGGCTGATGACCGGCAATTAACGAAGAGACACCAAGACATCT 239967
QY 767 TCTATGCTCGACCGGCGTGGGCCAGCGTCCGAACCCAGCGCGGACCGCGGCGGCA 826
Db 239966 ATTGG-----GGCGTCCGCAACGCGGCGCGGTGGCGGAGACCCATCCGCGC 239916
QY 827 CGCTGTATGACCAACACCGCGCTTTCGGGCGCTCCGACACAGGCGAGATTGTGCG 886
Db 239915 ACAATCTTACACTCGTGGGTGCTCGGCTCGATCCGAACAAGCGCAAGATCAAGACT 239856
QY 887 GTACCAAGACCTGCGCGCGCAACTGG3GCCAAGA 923
Db 239855 ATCAACAGTACCAACAGACGACTCTGCGGACTGGGA 239819

```

```

RESULT 9
AP005936/c 306600 bp DNA linear BCT 12-SEP-2003
LOCUS Bradyrhizobium japonicum USDA 110 DNA, complete genome, section
DEFINITION 2/31.
ACCESSION AP005936 BA000040
VERSION AP005936.1 GI:27348543
SOURCE
ORGANISM Bradyrhizobium japonicum USDA 110
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiaceae; Bradyrhizobium.
REFERENCE 1

```

AUTHORS Kaneke,T., Nakamura,Y., Sato,S., Minamisawa,K., Uchiumi,T., Sasamoto,S., Matanabe,A., Igesawa,K., Iriguchi,M., Kawashima,K., Kohara,M., Matsumoto,M., Shimo,S., Tsuruoka,H., Wada,T., Yamada,M. and Tabata,S.

TITLE Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110

JOURNAL DNA Res. 9 (6), 189-197 (2002)

MEDLINE 22484998

PUBMED 12597275

REFERENCE 2

AUTHORS Kaneke,T., Nakamura,Y., Sato,S., Minamisawa,K., Uchiumi,T., Sasamoto,S., Matanabe,A., Igesawa,K., Iriguchi,M., Kawashima,K., Kohara,M., Matsumoto,M., Shimo,S., Tsuruoka,H., Wada,T., Yamada,M. and Tabata,S.

TITLE Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110 (supplement)

JOURNAL DNA Res. 9 (6), 225-256 (2002)

MEDLINE 22485002

PUBMED 12597279

REFERENCE 3 (bases 1 to 300600)

AUTHORS Kaneke,T.

TITLE Direct Submission

JOURNAL Submitted (20-NOV-2002) Takakazu Kaneke, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research; 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0612, Japan

URL:http://www.kazusa.or.jp/rhizobase/,

URL:http://www.kazusa.or.jp/rhizobase/,

Tel:81-438-52-3935 (ex.2338), Fax:81-438-52-3934)

Location/Qualifiers

1. 300600

/organism="Bradyrhizobium japonicum USDA 110"

/mol\_type="genomic DNA"

/script="USDA110"

/db\_xref="taxon:224911"

/complement(28..309)

/gene="bgl0296"

/complement(28..309)

/gene="bgl0296"

/note="ORF ID:bgl0296"

/unknown\_protein"

/codon\_start=1

/transl\_table=11

/protein\_id="BAC45561.1"

/db\_xref="GI:27348544"

/translation="MRLIPVGVCAVCVALSGTASARGDGHSMASGALGTSAAAP GMSLGTALPSDSGGRATKXPLGTPTDIDDLRLKMLEGSLCRGC"

1017..2390

/gene="bglr0297"

/complement(1017..2390)

/gene="bglr0297"

/note="ORF ID:bglr0297"

/hypothetical\_protein"

/codon\_start=1

/transl\_table=11

/protein\_id="BAC45562.1"

/db\_xref="GI:27348545"

/translation="MHAQRPQYASLPGRGRTRRRVYVNPAPAMRRHLLITGRTY SLSGNAMRLGKLVISCAVLAVGSASLEPOYDASDTEIRIGNIMPYSGPASVYIG KMSAYFERININDGNGKINPISYDVASSSKAVEORRLINVEDDEVIALPAPGTA SNAALOKVNNHRRHVPOLPVITGASHMADDEHPMTIGMLPSRTREARVYADYLLKHPH GARGLIHQNDPFGRDYLLGLKDVLQDKAQMVASLPIEVESEPTIDSQVAVKSAAP DLFIDLATPKFAAQATRKLAEDWPHIHVTVSVSGVALXEPAGLIDNAKGLYSAGQ MDVTDPQMSHPGMOHFRAPMAKYPEADRSSEGPLVAFNMSALAVEVIRCGDGLTH ERIMVQVADLDELFAVYIPGIRIKITSPDFHPIEOLQMRFTGESWEFGPLDGRAD QATPLPD"

2353..2988

/gene="bglr0298"

/complement(2353..2988)

/gene="bglr0298"

/note="ORF ID:bglr0298"

/unknown\_protein"

/codon\_start=1

/transl\_table=11

/protein\_id="BAC45563.1"

/db\_xref="GI:27348546"

/translation="MDALTRRRPCPTRTSLMDPESMEVNPARELRKTKMFVNST LRIRAEHVSRAQGVRRRLFKEMDDYEAALVGRIRGSHDCLDRDSRQREAYO CRGICVGSASRGILQARLRPCSRRCVRRVPLAAAMRAHIGLISRTSSPPDAMWK LIVSTFRILTSRKEGRRLTRPPRLPAPHVSNTRHCSGGLATITIGEVSK"

3261..6173

/gene="bglr0299"

/complement(3261..6173)

/gene="bglr0299"

/note="ORF ID:bglr0299"

/product="transcriptional regulator"

/transl\_table=11

/protein\_id="BAC45564.1"

/db\_xref="GI:27348547"

/translation="MAGPLMSSPVDRRTYFEGHMLIDCORRELRNNNAVPIGRAREI LEKLVSGAGLITKDDIGSVMSGLFVEDNTLGVHISAVRKALGSDRNLLKTVSGGY TLGATWRSGLHCHPAAAPPTIPADALFNSNLPISIRAPILGRDLSLELSDVLSRAY TLVPGSIGCTKTLAIELARSVARSNDVSAVIELALQCALAASCTARALASNDK EYTPASIAQAIIGTRILLVLNDCEHYTEAAQASAVLHFCENVYLATISRSGLRIDG EYVAAPALSVDPDDMHAPDLIRASAVOLFVARTALRASFPVAGEIELOKIALICR LIDGIPLALEFAAARAANLGVDAVLRLDKRFLENGRSDOLPRHQTLEATLDWSYDL LPABEOLICOLAPPAGFTLIDASAAMDITLSRSDVETLEFVLVAKSIVSDPSEEG KRLLESTRAYALEKLADGLVEQAARQAVFROIILPGLDASPAPDISPRAOBIG NYHRAIDMASPSPGDTALGIELTSGFVPVMOLISFEGSTRIEHLAHSRSGFTCP KLRQULYVNGRPLINTTSGADMKRALSIGALAEISDLIELKAWTILSYNINS GQYAAADVGERYLELALRTGNPANAVGRRLIGAAHFGAOEVARELITLSLSDSA HGLSGGTLNMLFLNOSVLAKMLARVLLQGIQASRLAESCLODAREDKLAIA YALRNACPIPALMTDHLAADAQAISSILBLEYREGIAFWTSWNSCLKQGLVYOGHD DGIALLRSGKARTENGLMKRNEPFGSLAEGEAGAAQALAAVEALSLISGRQD LMCIALDRIKAEIILANGSELSRESELFAEGLSVARQOCSFYELKAAAMARIMA KSNRQGTALNDIVGSPVSLAPQOETDLALPFAARLALPIGISLSTYGEKSPGOTCH"

complement(6919..8190)

/gene="ragD"

/complement(6919..8190)

/gene="ragD"

/note="ORF ID:bll0300"

/codon\_start=1

/transl\_table=11

/product="RagD protein"

/protein\_id="BAC45565.1"

/db\_xref="GI:27348548"

/translation="MSDVERERSDDEADNRPGAESRLIYELPGKVARSKRRYGGPLG GYVILLAGLIGIGRWNYEAREVAVAEOVTNPEVRAVAQVQSGMLKXTLAT TTAPEANITPASTGYIEKRYVDIGRVKKGDLLVETLPELQQAQALQALTAQOA ALQOASARELADVTNSRDSNLYKQGMILTAQOGDNDRLTLAGQQAAGVQASVIAQO AQRIVLQOESYORVAVAPPGYVTOGNVNGSLYQAGSTMTMMSNVTIRQVFPDQ DEAPGAVGVDSDIRYVEIPGRTPAGVTYRATLALPGSRTLLITEIDVNPDSGLSPG IYCTVELSLIPRRTPTTIPSDALVPQNGLHAVVANGRVHFOQVLSRDFGTTVEVR DGVPQGVVLANFAVNLAEQSKTVKRSVS"

complement(8187..11369)

/gene="ragC"

/complement(8187..11369)

/gene="ragC"

/note="ORF ID:bll0301"

/codon\_start=1

/transl\_table=11

/product="cation efflux protein"

/protein\_id="BAC45566.1"

/db\_xref="GI:27348549"

/translation="MGIYRPALELPHRTFYVLAALLIFLGIAIRSPMDIIFPRLIPV VLIWVSTGSLSTEMEORVSYSQISANVTGIRKITEQTLNGLSVQKTIYOPDNL DLAISQVSATVNAIRALMPGLOPIIIVQFNASSVVLQGLSLSNLSQKQILQDFIY RVRQOALPVGVTLPPTAGKRYQIWDIDPNKLISRGTLPLDIYNAVNTQNLTLPTG TTKIGDQVYRTNATPATIODLNMIPVKFANGATFLKDVQVDRGAQVOQVIVED GHRVAIVSVKNGNASTLAVNGVKALISIRASAPAGKINELPQSLIFVHSVNGV LRCGALPAGLTALMILVPLGWSRSTLYVLSIPLMSSLVLYLFGELTNMTLGLL ALAVGTLVDSSTYTITENTRILWTEBGMPLSEATLHGAATAIPTVSTLAISGVPTSV ALVGGPKYLPPLGAVVPAMLASTGSRITLPTITGLLKKERRHVGGBRSNIFA RASAPFERGFENLRDGYSDLLTLRLRVIVPVAVLVALGATKRVVYGRDYPILD GGOIOLHVRAPAGTRIERTEALFOAVEDKIREVIRPBDALVDNIGLARPYNLAFT DGSITIGVNDGITLIVSLKDGHKPTADYIRKLRLRPLPAPAFEDTFFVFOADIVVOILNFG

LAIPADIVRTVYSGNNILAVAKELQOEWLTPGVNDHLOOEVDPAFASIDIPRRAO  
GLNAISVAVININIVINIVSSASQENKMTDPTSGIYIYALVOPEKXANSINSGTVEP  
SASISAGCQYKTLNVAUPEKQYCPNTUNNOOPDYDASISQVGDLSASADID  
KYTSLQKQLOPKNSLOYGQLOSMQOSRDIQIGLLPAILIYVILAVNTRIONFGD  
VYINALPITPOGIVYMLPVYGTITLNPSSLMGAIMVGVASANSILVLVFAEDBEDKCH  
SAPALASASGSTRIRPVLMTLNVAVIOMT PMA IGGGEONNOALARAIVIGLLPATPT  
LILVLVPLIFPMURKZDNGKRHHGVCEPDEE

gene	complement (11449. .12372)
	/gene="rpoH3"
CDS	complement (11449. .12372)

CDS

```

/gene="rpoH3"
/note="ORF_ID:b110302"
/codon_start=1
/transl_table=11
/product="RNA polymerase subunit"
/protein_id="BAC45567.1"
/db_xref="GI:27348550"

```

/transhettvnsasvnaagaavapfnsavsaalrreyllebgqo  
 qlarshettvtdrgadavlvtshllraaklarqkglpmdvllleavlgsviavsf  
 epbgafstyaivmiikaahlliesmsvltvgtaakqlfblreslkrasvff  
 sglpvdvaeliagldvltvarevtemdarngmslnarageestelbalvngavd  
 aeytladheqetervaralaloglaaarrhryfearlrltecpvllldqlarlsisssr  
 /rrolreiafakvraawlaaodprravcnv/

gene	complement (12433. .13857)
	/gene="ragB"
CDS	complement (12433. .13857)

CDS

```

/gene="ragB"
/note="ORF_ID:b110303"
/codon_start=1
/transl_table=1
/product="two-component sensor histidine kinase"
/protein_id="BAC45568.1"

```

```

# /db_xref="GI:27348551"
#/translation="MHOSILRSTPIMLAVALAAAFALFVYGLFAFYIKWDYLVAR
SDGATIRHIFIDLPARISIAIHLQDSRGVYGLDAGTIRLAAGIDHLP,PE
LGDIAAVQVRGLPEKPAHVAVAVRRLDGGD,WMERNDEHRETSAYQGL
ALGILPAST,CILAGMS,SRAPRVEPNPVSIVAGEL,RELP,PEGSDP,PARIA,
GIWYMDIEMETIMNMLAGCGD,IAADDET,ITR,IALBERGTHAAKIDLEOETIR,
ALAIQDOS,LAIVALLATGIRIBNNRMAFGEVALDELIRSCVDEYFIDADR,LAAG
VIVERNQWAGDELD,FEAAVNLVDNAVAF,TSGRGVR,LA,EAEGDILRVSDTSG
IRROBERAAVRRIRYSKDKNNTGSLVLAIVKRLHGRLIVGPAPGRGLEITLAW
TGRRDKAATRSRSTLHPAADLNG"
# complement (13861)

```

ସ୍ତ୍ରୀ

Query Match	6.3%	Score 110.2;	DB 1;	Length 300600;
Best Local Similarity	46.7%;	Pred. No. 1.8e-10;		
Matches 736;	Conservative 0;	Mismatches 763;	Indels 78;	Gaps 9;

QY	94	CTGCTGCAGAACCCGCCCGCTGGTGGATTTAACTACGGCCGCACACAGAAAATAT	153
Db	50383	CTCAAGCGCGAATCCGACTCGAGCGACTGGCGCCCTTCCCGCCACGACTAGGGACAACG	5032
QY	154	CGCCACTCGCCCTCGACCCAGATCACTGCCGACAACGTTGGTCACTGGCAACTGCTCGG	213
Db	50323	CGCTTCAGCCCGCTGGAAGCAGATTAACAACCGAGAACCCCGCAAGCTGACGCTGCTAT	5026
QY	214	GCCCGCGGAGTGAAGGCG---GGGCGCGTACAGGTCAAGCGCGATTCATGATGGCGTG	270
Db	50263	TCGTTTCGGCTCGGCTCGCTCGCTCCGTTCCAGACGAACTCTGGCGATCGTCTGTCGTAAACG	5020
QY	271	ATGTAATCGCAAAACCCCGGTGATGTG-----TCAAGGCGCTGGATGCGCAAAACGGC	324
Db	50203	CTCTACGTTTCGACACTCTCTGGGGCCGCAAAATACGTCTACGCGCTCATGCGCCACCGGC	5014
QY	325	GATTGTGATCTGGAGAACACCGCGCCCAACTGCGCCCGCTGCCACGCTAAACGCCCAAGC	384
Db	50143	GTCGGCAATGACCTTACGAACCCGATATTTCCGAGAGAGCGTCAATACGCTGTGTC	5008
QY	385	GACCG---CAAGCGCGCGCTGCGCCCTTATGGGACAGAGCTCTATTTCAGCTATGGGAC	441
Db	50083	GACGTGAACCAATCCGCGCGGTCTCTATTAGCCGACGGCAAGATCTTTCGTCGGCGCGCTGAC	5002
QY	442	AACCATCTGATCGCGCTGAGATATGACAGGGCCAGGCTGTAATCGATGTGCAACTGGA	501

Db	50023	GGCAAGCTGACGGCGCTCGATGCGCGGACCGGCAAGG	-GGCTCTGACCGGCAAGGTGGT	499655
Qy	502	TCGGGCGCAAGACGGCTTTGACCAAGTAACACCA	CGGGGCCGATTTGTGCGCAATGGCGTCAATC	561
Db	49964	CGATTACAAACAGGGGTTT	-CGGTATCACTCCGCCCGCTCGTCTGGCGCACAAGGTC	499070
Qy	562	GTGGCGGGGTTCCACCTGGCCCAATTTGGCCCTATAG	NAAGTCTTATCTCGGGGGCAAGATTCC	621
Db	49906	ATCACCGGTTTTCGGCGGGCGGCGAGTACGGGTG	CGCGGCTTCGTTCCAGGCTTTGCAATC	498477
Qy	622	GCGACGGGTAGGAGAGCTGTGGCCCAACACTT	TATTCGCGAGCCGGCGCAAGAGGGGTGAC	681
Db	49846	AAACACGGGCAAGCAGCTGTGGCAGACTTACAC	CGTTCCGTTCGGATGAGCCCGCGGG	4978
Qy	682	GAGACTTTGGGGCAATGATTTTCGAGGGCGCGT	GATACCGGCGTCTGGGGTCAGATCACC	741
Db	49786	GACACCTGGAAAGGCGCACTCCGCCACAGATGG	GCGGCGTTCGGGCTGTGGCTTCG	49722
Qy	742	TATGATCCCGTGAAGCACTTGTGTTCTATAGCT	TCGACCGCGCGCCAGCGTCCGAA	801
Db	49726	TAGATATCCGAAGACCAACAGCTCTATTTGGGG	ATCCAGCAATCCGCGTCTCCGGAACAG	4966
Qy	802	ACCCAGCGCGGACGCGCGGGCGACGCTGTATG	GCACCAACCCGCTT-----	850
Db	49666	GCGGTATGCTCGACCGGCGACCGCAATTTCCG	CAACTGACCAACTCTTACACCGCCTCG	496070
Qy	851	-----TTGGCGGTGGGTCCTCCGACAGGGCGA	AGATTTGTGGCGTTCACAGACCGTCGCGCG	906
Db	49606	ACCGTGGCGTTTCGATCCCAACACCGGCGAAG	ATAAATGCGACATCCAGACACCGCGCGC	495477
Qy	907	GACAACTGGAGCCCAAGAAATGCAAGTTGAGAT	ATGTCGCCAACCTCATGTGCAACCC	966
Db	49546	GATGCGTGGGACTATAGCGGCGTCAACGAAG	CGGTGCTCGCGGATCTGAAGATCGGCGCG	494870
Qy	967	TCGGCGCAAGATGAGAGGCTTCGGCGCCATCA	ACCCCAATGCGCGCACGGGCGAGCGCGCT	1026
Db	49486	AGCAACCGTGGCGCGCGTGAAT-----	GAAAGCGGATGCGCAACCGC	49448
Qy	1027	GTCGTGACGGGTACGCGCTTGTSCAAGACCGG	CACAGATGTGCTGTTTTATTCGGCGTCCGGC	1086
Db	49447	TACTTTCTGTCGCCCAATCGCGAGACCGGCA	AGGTGCTTTTCGGCGGAGAAATACGTTTC	49388
Qy	1087	GAAATTCGTGGCGGCGTGAATACCAACTTAC	CACTATATGTCGCTCGATGACAGAGACC	1146
Db	49387	-----TCCAACTGGGCGCAAGAAATGGGAG	ACTCGCCACGATGCGC	49349
Qy	1147	GCGCTTGTGACGGTGAACAGAGATGGGGGTG	CTGAAAGACGTGGATGATATATACGTC	1206
Db	49348	GCGGTGCAAGGATCCCGACAAAGCGTCCCGG	CGCGCAACATTCGCGC-----AAGGACATC	49295
Qy	1207	TGGCCGACCTTCTCTGGGATGGGCGCGACTG	ATGCTGACGCGCACTGAACCCGAGCACCGCG	1266
Db	49294	TGCCCGAAGCTGATCGGCGGCAAGAACTGG	CAGCGATGTCGTTCAACCCGAGACCGCG	49235
Qy	1267	ATTTACTTTTGGCGCGTGAACAAATGCTGTG	CTACAGATTTATGCGCGTTATCAAGAGTTT	1326
Db	49234	CTCGTCTACATCCCGTCCAAACATGTCTCA	TGAGATGTGCGTCTCCGACGTGCTAC	49175
Qy	1327	AGCGCGTCAAGCTCTATTAACACACAGCGG	ACCGCAAAATCTCGCGCGGGCTTTGAAAT	1386
Db	49174	AAAGCGGCGTGTGTTTATCTCGCGCGCGAA	TTCCGACCAAGAGAGCCCGCGGTTTC	49115
Qy	1387	ATGGGCGCGATCGACGCGATGATTAACAGC	ACCGGCGCACTTGTGTGCGCGAGCGC	1446
Db	49114	CTCGGCGAGCTCTGTGCTGTGGATCCGAT	TCGCGCAGAAAGATCTGTGCTGATCAAGAA	49055
Qy	1447	CTGCGCGGCAACATACCTGCGCGTTTTGTG	AGCGAGCGGATGTGTTTCAACGCGCGG	1506
Db	49054	GACCTGCGTTTCAAGCGCGGCAAGCTTAC	CAACCGGCGGTGCGCTGTGTGTTTCGCGCAAC	48995
Qy	1507	ACCGACCGGTAATTTCCGTGCGCTTCAAGC	AGAAACCGGCGAGACTTTGTGACAGCGCCT	1566

Db 48994 ATCCATGCGGATTTCCGCCCATTCATGCGAAAGTCCGCCAAGTGTTCGACGACGAT 48935  
Qy 1567 CTTCGACGCGTCCGACGCGGCGACGCGATCACTAGAGTTGACGCGCGTGCATATATC 1626  
Db 48934 CTGCGCTCCGCGATCCGCCAGATCCGATGACCTACACAGTGCAGCGCAAGCATATGTC 48875  
Qy 1627 GCCATGCGTCCGCGCGG 1643  
Db 48874 GCCATGCGTCCGCGCG 48858

RESULT 10  
AF277373 2097 bp DNA linear BCT 07-MAR-2001  
LOCUS Ralstonia eutropha tetrahydrofuryl alcohol dehydrogenase gene,  
DEFINITION complete cds.  
ACCESSION AF277373  
VERSION AF277373.1 GI:9255864  
KEYWORDS  
SOURCE Maurexia eutropha  
ORGANISM Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
Burkholderiaceae; Maurexia.  
REFERENCE 1 (bases 1 to 2097)  
AUTHORS Zarrt,G., Schraeder,T. and Andreesen,J.R.  
TITLE Catalytic and molecular properties of the quinohemoprotein  
tetrahydrofuryl alcohol dehydrogenase from Ralstonia eutropha  
strain Bo  
JOURNAL J. Bacteriol. 183 (6), 1954-1960 (2001)  
MEDLINE 2112557  
PubMed 11222593  
REFERENCE 2 (bases 1 to 2097)  
AUTHORS Zarrt,G., Schraeder,T. and Andreesen,J.  
TITLE Direct Submision  
JOURNAL Submitted (13-JUN-2000) Microbiology, University of Halle,  
Kurt-Moche-Str. 3, Halle 06120, Germany  
FEATURES  
source  
1..2097  
Location/Qualifiers  
/organism="Maurexia eutropha"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:510"  
1..2097  
/codon\_start=1  
/transl\_table=11  
/product="tetrahydrofuryl alcohol dehydrogenase"  
/protein\_id="AAF8635.1"  
/db\_xref="GI:9255865"  
/translation="MHSNKLRLVLCVMAAASVALPAPMAFGANAARVDAALRANEA  
GPNMPSVGLDYAEFRFSKLEQVNAAGNVLGLAWSYDLSEIRGVEATPLVDGMYV  
SAFMSVVAHLDARTGKRLMTYDPQVPRDAVYGCDDVNRGVALYOGAFVGAEPGRU  
VAIDATGKTVWQDTIVDRSKSYITIGAPRYNGKVIINGAGATGVRGTTATDAE  
TEKQQRWYTVTPDPAPRPENBAMAAATWDPGSKYWIINGGGVMTWMTMAFDEPLNL  
MYITGNAGPWSKILSPKGDNLVAASVALNPDTGEVHWYQTPGDNDYSTD  
IIADLKIDGQPKRVLIIHAPKNGFEVIDRTNGKFIISAKNPVDVMAGSYKNGPVE  
TPQADTSKPADVDPGPGAHNMSHFELKGLAFIPQAHVPLTLDNKKMVENQKD  
SPAHSGVMNLCMLVNAAPRSPKPMGRILAMDPLAOKVYMRDLAGMNGGLATAG  
NVTFQGTADGRVYAHATGEEKLMQAFPTSSGVAAAPVTLIDGRYVSVAQWKGVYG  
LSGRGSDQAPGKIVTVFLNGKTPMFPASAPLAPULSGVKDPAKVDQKLTYYNAC  
VFQGVGVADKGGKINPLAYVGAPMIEHLDKLFGPFVQKMPDTKLTAEDEVKLI  
QAFICGTADAVPRPK"

ORIGIN  
Query Match 6.3%; Score 110; DB 1; Length 2097;  
Best Local Similarity 48.8%; Pred. No. 3.4e-10;  
Matches 450; Conservative 0; Mismatches 430; Indels 42; Gaps 4;

Qy 119 AATGATTACTACGCGCCGACCAAGAAATATATGCGCATCGCCCTGACCCAGATCA 178  
Db 143 ACTGGCCAGCTACGCGCTGATTACCGCGAGACCGGCTTCAGCAAGCTCGACGAGTCA 202  
Qy 179 CTGCGCAACAGCTTGTCTGATTCATCTGTCTGGGCGCGCGGAGTGA---GGGCGGG 235  
Db 203 ACGCGGCGCAACGTCAGGAACTTGCGGCTGGCGTGTCTTACGACCTTGGAATGACGCGCG 262

Qy 236 CCTTACAGGTCAAGCCGATATCCATGATGAGCGTTATGTAATCTGGAAACCCCGGATG 295  
Db 263 GCGTGAGGCGACGCGCTGCTGTGTGACGCGCTGATGTAATGTCCTCGCGCGTGAAGG 322  
Qy 296 TGATCCAGGCGCTGATGTCGCAAAACAGCGCATCTGATCTGGGAACAACCGCGCAACTGC 355  
Db 323 TGTGATGATGCAATGACAGCCCGACCGGAGAGCGCTGTGGACCTACAGCCGCAAGTGC 382  
Qy 356 CCGCGGTGCGCAAGCTTAAAGCCCAAGGCGAC--CGCAAGCGCGCGCTTCCCTTTAG 412  
Db 383 CGCGCGACAGGCGCTTACAAAGGCGCTGTCTGCGACGTGATCAATCGCGCGTGGCGCTTACC 442  
Qy 413 GCAGAGCTCTTATTTAGCTCATGAGGACACACCATCTGATTCGCGCTGATATGAAACGG 472  
Db 443 AGGCGAAGTCTTCTGCGCGCGCTTGAACGCGCGCTGTGGATCATGACGAGCAACCG 502  
Qy 473 GCCAGTCTGATTCGATGTGAACTGTGATCGGCGCAAGCGGCTTGACCATGAACCA 532  
Db 503 GCAAGAAAGTCTGGGAGCAGAGACACATCGTCACCGCAGCAAGTCTTACACCATACCG 562  
Qy 533 CGGGCGGATTTGCGCAATGCGGTGATCGTCGCGGTTTCACTGCAATATTCGCCCT 592  
Db 563 GCGCGCGCGCTTACACAGGCAAGTATCATCGCAACGCGCGCGGATATAGCG 622  
Qy 593 ATGATGCTTTATCTCGGCGCACGATTCGCGACGCGGTAGAGCTGTGCGCAACCAT 652  
Db 623 TGGCGGCTATATCACCGCTTACGATCGCAGACCGGCAAGCAATGCGCTGTGATCA 682  
Qy 653 TTATCCCGACCGCGGCAAGAGGTGACGAGCTTGGGGCA----- 694  
Db 683 CGGTGCGGCGACCGCGCGACCATTCGAGAACGACGACGAAAGCGCGCGCA 742  
Qy 695 -----ATGATTTGAGCGCGCTGTGATGACCGGCGTCTGGGGTCAAG-----A 736  
Db 743 CTTGGATTCGAGGGAATGATGATCAACGCGCGCGCGGACGATGTGAAACCA 802  
Qy 737 TCACCTATGATCCCGTACGACAACTTGTGTCTATAGCTGACACCGCGTGGCGCGACGT 736  
Db 803 TGGGTTGATCCCGACCTCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 862  
Qy 797 CCGAAACCCAGCGCGGACCGCGGCGGCGACGCTGTATGCGCAACACCGCTTTGCGG 856  
Db 863 GCGCGACGCTCGCGAGCGCGGCGGAGGATGATGATGATGATGATGATGATGATGATGATGAT 922  
Qy 857 TGGCTCCGACAGCGGCGGATGTTGTCTGCGCTACCAAGACCTCGCGCGGATGATGATGATGAT 916  
Db 923 TGAACCCCGACACCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 982  
Qy 917 ACCAAGATGACGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 976  
Db 983 ACTACACTGACCGCAGACATCATCTGCGACCTGAAGATGACGCGCGCGCGCA 1042  
Qy 977 TGGAGGCTTGGCGCGCATCA 998  
Db 1043 AGTGATCTCGATGCGGCCAA 1064

RESULT 11  
LOCUS CTGHEBH 2959 bp DNA linear BCT 12-APR-1996  
DEFINITION C.testosteroni qheh gene.  
ACCESSION X81880  
VERSION X81880.1 GI:663195  
KEYWORDS qheh gene; Quinohemoprotein ethanol dehydrogenase.  
SOURCE Comamonas testosteroni  
ORGANISM Comamonas testosteroni; Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
Comamonadaceae; Comamonas.  
REFERENCE 1 (bases 1 to 2959)  
AUTHORS Stoorvogel,J., Kraayveld,D.E., Van Sluis,C.A., Jongejan,J.A., De  
Vries,S. and Duine,J.A.  
TITLE Characterization of the gene encoding quinohemoprotein ethanol



dehydrogenase of *Comamonas testosteroni*  
 Eur. J. Biochem. 235 (3), 690-698 (1996)  
 MEDLINE  
 PUBMED  
 56184549  
 8654419  
 2 (bases 1 to 2959)  
 Stoorvogel, J.  
 Direct Submission  
 Submitted (16-JAN-1995) J. Stoorvogel, Department of Microbiology  
 and Enzymology, Delft University of Technology, Julianalaan 67,  
 2628 BC Delft, NETHERLANDS

FEATURES  
 source  
 1. .2959  
 /organism="Comamonas testosteroni"  
 /mol\_type="genomic DNA"  
 /strain="ATCC 15667"  
 /db\_xref="taxon:285"  
 /cell\_line="E.coli JM109"  
 350. .354  
 361. .2487  
 /gene="qhedh"  
 361. .2487  
 /gene="qhedh"  
 /codon\_start=1  
 /transl\_table=11  
 /product="quinohaemoprotein ethanol dehydrogenase"  
 /protein\_id="CA57464.1"  
 /db\_xref="GI:661196"  
 /db\_xref="GOA:Q46444"  
 /db\_xref="Sw:SS-Prot:Q46444"  
 /translation="MERLIDNSHGFGRNVTLLACGSAAPFATGPAQAQAAVOR  
 VGGDFIRANAARTPDWPTIGVDYAEIRYSRLDQINAAVKDGLAMSVNLESTRGVEA  
 TPVVDGIIYVSAWSVVAHIDRTGNRIWTDPOIDRGTGGCCDVNRGVNLGVEA  
 KYVVGMDRLILALDAATGEVWHONTPEGOKSLITITGAPRVKXVILIGRGAEYG  
 VGGYITADAEGERKMEVSGDPSKPEDESMRAARTWDPGKMEAGCGGTMM  
 DSWTPEALNTWTVGTNGSPMSHKVRSFGGDNILASIVLADPDGTGYKWHVETP  
 GNMWDYTSQPMILADIKTAGKPRKTIILHAPKGFPTVLDRTNGKFIKSNVFNMA  
 SGTRKGRPTIGLAARDGSKPDVAVGPTGANNHMSFNPTGLVTLPAQNVFNLM  
 DDKMEFNQAGPKPQSGTGMNTAKFNAPKSPKPFGLLMDPAQKAAVSVHS  
 PNVGGLTTPAGNVFQGTADGRILVAHAATGKLEWAPRTGTVAAASTYWDGQYV  
 SVAVGWGVGLAAERATERGQGTIVTVFVAGKAMPFVAGRTQLQGVYDPAKY  
 EAGTMLVAVNVCVCHGPGVDGNGTPNLGVYDASYIELNPFVFKGPAMVGMWDPF  
 GLTSGDDVESLAKFIQGTADATRPKE"

sig\_peptide  
 361. .453  
 /gene="qhedh"  
 454. .2484  
 /gene="qhedh"  
 /product="unnamed"

## ORIGIN

Query Match 5.8%; Score 101.2; DB 1; Length 2959;  
 Best Local Similarity 47.5%; Pred. No. 1.3e-08;  
 Matches 439; Conservative 0; Mismatches 443; Indels 42; Gaps 3;

QY 117 TGAATGATTAATACGCGCGCAACCAAGAAAATCTATCGCACTGCGCCCTACCCAGAT 176  
 DB 534 TGACTGCCCCACATGCGCGTGAATATCCGAGACCCCTACAGCGCTGATCAAGAT 593  
 QY 177 CACTGCCGCAACGTTGTGATGTCAGTCTGTGGGCCCGCGGAGTAGAG--GCAGG 233  
 DB 594 CATATGCCGCAACGTAAGAGACCTGGCGCTGGCATGCTGTCAACCTCGAGTCCACGCG 653  
 QY 234 GCGCGTAAAGGTACAGCCGATGATCATGATGCGGATGATCTGGCAAAACCCGGTGA 293  
 DB 654 CCGCGTGAAGGCCACGCTGTCTGTGTGACGCGATCATGATGTACGCGCTCATGAG 713  
 QY 294 TGTGATCCAGGCGCTGATGCGCAACAGCGCATCTGATCTGAGAAACACCGCGCCACT 353  
 DB 714 CGTGTGATGTCATGACGACCCGTAACCGCAACAGAGATCTGACCTATGACCCGAGAT 773  
 QY 354 GCGCGCGGTGCGACCGCTAAAG--CCCAAGCGGACCCCAAGCGCGGTGCGCCCTTAA 410  
 DB 774 CGACCGGACGACCGGCTTCAAGGGGTGTGCGACGTCGTCAACCGCGGCGTGGCTGTG 833

QY 411 CGGACAGACGCTTATTCAGCTCATGAGCAACACATGATGCGCTGGATATGAGAC 470  
 DB 834 GAAGGGCAAGGTCTATGTGGGGCGTGGAGATGCGCTGATGCGCTGATGCGCGAC 893  
 QY 471 GGGCAGGTCTGATTCGATGTCGACGTGATCGGGCGAAGACCGCTTGACATGACAC 530  
 DB 894 CGGCAAGAGGTCTGGACCCAAATATCTTCGAGGGGCAAGAGGGGTGTGCTACATCAC 953  
 QY 531 CACGGGCGCATTTGCGCAATGCGCTCATGCTGTGGCGGTTCACCTGCCAATTTGGC 590  
 DB 954 CGGCGCCCGCGCTGTTCAGGGCAAGTCATGTCGCAAAACGCGCGCGCAATTTGG 1013  
 QY 591 CTATGATGCTTTATCTCGGGGCGACGATTCGCGACCGGTGAGAGGTGTGCGCA---- 646  
 DB 1014 CGTGGCGGCTATATCACTGCTTATGACCGCGAACCAGGGAGGGGAAATGGCGCTGTT 1073  
 QY 647 -----ACCACTTATCCGCGACCGGCGGAGA 674  
 DB 1074 CAGCGTACCCGCGCATCCCTCCAAACCCCTTCGAGAGACGATGCAAGCGCGCGCCAG 1133  
 QY 675 GGGTGAAGAGCTTGGGGCAATGATTTGAGGGCGCTGATGACCGCGCTGCGGTCA 734  
 DB 1134 GACCTGGACCCCGCGCAATATGTGGAGGCCGCGCGCGCGCACATGTGGACAG 1193  
 QY 735 GATCACCTATGATCCGTCGACGAACTTGTATTGCTATGCTGACCGCGCTGGGCCAGC 794  
 DB 1194 CATGACCTTGATGCTCGCACTCAACCATGATGCTGGGACCGGGCAATGGATGCGCTTG 1253  
 QY 795 GTCCGAAACCGAGCGCGGACGCGCGGCGGACGCTGTATGCGCAACACCGCTTTGC 854  
 DB 1254 GTCCGACAGGTTGGGACGCGCCCAAGGGCGGAGACACCTGTACTGCTTCATGCGGC 1313  
 QY 855 GATCGTCCCGACGCGCGCGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 914  
 DB 1314 CCTGATCCGACACCGGCAAAATGAAAGTGCATATCAAGAAACACCGGCGCAACTG 1373  
 QY 915 GGACCAAGATGACAGCTTCGATGATGATGTCGCCAAGTCGATGTGCAACCTTGGCGCA 974  
 DB 1374 GGAATTAACCTCCACCCAGGCCATGATCTGGCGCATCAAGATCGCGGCAAGCGCG 1433  
 QY 975 GATGAGAGGTGCGCGGCATCA 998  
 DB 1434 CAGGTACTACTGATGCGCCCA 1457

RESULT 12  
 AF176640  
 LOCUS  
 DEFINITION  
 Pseudomonas stutzeri sigma54-dependent transcriptional regulator  
 AdhR (adhR), alcohol dehydrogenase (adh), coenzyme PQQ synthetis  
 protein A (pqgA), and coenzyme PQQ synthetis protein B (pqgB)  
 genes, complete cds.  
 ACCESSION  
 AF176640  
 AF176640.1 GI:9957180  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Pseudomonas stutzeri  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.  
 REFERENCE  
 1 (bases 1 to 7097)  
 Chang, C.-H., Herrick, J.B., Okinaka, R.T., Brainard, J.B. and  
 Terwilliger, T.C.  
 Identification and characterization of genes activated by  
 2-chloroethanol in *Pseudomonas stutzeri* BC-2  
 Unpublished  
 2 (bases 1 to 7097)  
 Chang, C.-H., Herrick, J.B., Okinaka, R.T., Brainard, J.B. and  
 Terwilliger, T.C.  
 Direct Submission  
 Submitted (09-AUG-1999) Life Sciences Division, Los Alamos National  
 Laboratory, MS-M888, Los Alamos, NM 87545, USA  
 FEATRES  
 Location/Qualifiers  
 1. .7097  
 source



QY 1125 GATGCGCTGATGACGAGACCGGCGCTTGTGACGTTGAAACGAGATGCGGTGTAAGA 1184  
Db 4351 GGGCGAGAGAGTGTGATCTGGCCACCGGCGCCCGGTGAGGTCTCCGGCTCGGCTACGA 4410  
QY 1185 GCTGAGCGTTGAATATGACGCTCTGCCGACCTTCTGGTGGGCGGACTGTGTGACG 1244  
Db 4411 GAAGGA---GCAGGTGGTGTATGTGGCGGAGCTCGTTGGGGCGGCAACATGGCACTCCAT 4467  
QY 1245 CGACATGAAACCGCGGACACCGGCACTTACTTCTTCCGCGCTGGAACAATGCTGTACATAT 1304  
Db 4468 GTGTCTCAACCCGCGACGCGGCTGATGATACCTCCCTACAGAAATCCCGGCGTCTTA 4527  
QY 1305 TATGCGCGTTGATCAAGAGTTTACCGCGCTCGACGTCTATAACACGACGCGCAAA 1364  
Db 4528 TCGCAAGAGAGGGGCGCACGTTCAAAAAGATCGACGCGTTGAAACACCGGATACCGGCTTACG 4587  
QY 1365 ACTGCGCGCGGCGCTTTGAAAAATATGGGCGCGCATGACGCGATTTGATATACACCGGGCG 1424  
Db 4588 CGACACCCCAAGATACCCCGCATGCGGTGACGCGCGCGCTGCTGGACCCGCT 4647  
QY 1425 CACCTGTGTGTCGCGCGGAGCGCCCTGGCGGCACTACT-----CGCGCGTTTGTTC 1475  
Db 4648 ACCGCAACCGGAGAGGCTGGGCGCTGCCGCACTACTTCTACTGACAGGCGGCACTTGAG 4707  
QY 1476 GACGGCAGGGCGGTGTGTGTTCACGCGCGGACCGACCGCTATTTCGTGCGCTCAGCCA 1535  
Db 4708 CACCGCGCGCAACCTGTGTTCACGCGGACCGCGACGCGCACTGACGCTACTCGAG 4767  
QY 1536 GGAACCGCGGAGACTTTGTGTGACGCGCGCTCTTGTGACGCTGCGACGCGGGAGGCGAT 1595  
Db 4768 GGCACAGGAGCAAGCGCTGTGAGACTTGCAGCGGACGCGGATGTCGCGCGCCCAT 4827  
QY 1596 CACCTACGATTTGACGCGCGCTGCAATATATCGCATGCTGTCGCGCGG 1643  
Db 4828 CACCTTCAGCTGACGCGGACGATACGTGCGGTGATAGCGCGGCTG 4875

RESULT 13  
AE016784  
LOCUS AE016784 302101 bp DNA linear BCT 12-DEC-2002  
DEFINITION Pseudomonas putida KT2440 section 11 of 21 of the complete genome.  
ACCESSION AE016784 AE015451  
VERSION AE016784.1 GI:26557028  
KEYWORDS Pseudomonas putida KT2440  
SOURCE Pseudomonas putida KT2440  
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
1 (bases 1 to 302101)  
Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,  
Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,  
Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,  
Nelson,W., White,O., Peterson,J., Kouri,H., Hance,I., Lee,P.,  
Holtapple,S., Scanlan,D., Tran,K., Moazzez,A., Utecherack,T.,  
Rizzo,M., Lee,K., Kosack,D., Moestl,D., Medler,H., Lamber,J.,  
Hohelsel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timms,K.,  
Duesterhoft,A., Tummler,B. and Fraser,C.  
Complete genome sequence and comparative analysis of the  
metabolically versatile Pseudomonas putida KT2440  
Environ. Microbiol. 4 (12), 799-808 (2002)  
2 (bases 1 to 302101)  
Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,  
Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,  
Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,  
Nelson,W., White,O., Peterson,J., Kouri,H., Hance,I., Lee,P.,  
Holtapple,S., Scanlan,D., Tran,K., Moazzez,A., Utecherack,T.,  
Rizzo,M., Lee,K., Kosack,D., Moestl,D., Medler,H., Lamber,J.,  
Hohelsel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timms,K.,  
Duesterhoft,A., Tummler,B. and Fraser,C.  
Direct Submission  
Submitted (05-NOV-2002) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA

FEATURES  
source  
Location/Qualifiers  
1..302101  
/organism="Pseudomonas putida KT2440"  
/mol\_type="genomic DNA"  
/strain="KT2440"  
/db\_xref="taxon:160488"  
200..1066  
/locus\_tag="P2639"  
200..1066  
/locus\_tag="P2639"  
/note="Identified by match to PFM protein family HMM  
PF00701"  
/codon\_start=1  
/transl\_table=1  
/product="aldehyde dehydrogenase, putative"  
/protein\_id="AA068247.1"  
/db\_xref="GI:24984217"  
/translacion="MSNFRGIWIALVTPEANEIDPFGLEKLYKKLIEDGAYGVC  
TTGGAALSKAEQALVDALVAVPEGRMWSGYNREILAFQASIQORDIGLLV  
PAPCYIRPSQAGIEAFPAIVADASVVIYVDIFRYGVIREBTLIVRHPILAV  
KDCGDEBTWALIEDGAYLAGEDEIQFNILCLGAGAGIISASAHYRADILYAMVO  
VDSGDMAARGTFFQLPWIKVAFAPENPAVKALQLOSLADELREPMQTCNTR  
DKLSVLCISGA"  
complement(11220..1678)  
/locus\_tag="P2640"  
complement(1220..1678)  
/locus\_tag="P2640"  
/note="similar to GB:M23326, GB:X15261, GB:X13954,  
PID:1049195, PID:2358067, PID:312412, PID:37046,  
PID:37048, PID:37298, PID:37324, and PID:540457;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=1  
/product="acetyltransferase, GNA1 family"  
/protein\_id="AA068248.1"  
/db\_xref="GI:24984218"  
/translacion="MOVRLTTPDDLPQNALCNDAPMOAVAPTLASGIEFTQVSAP  
QAPAFRQGDNLMLACFVEGALILIKGRHILAMLPFAPGLOROSIGKIMMALE  
HASAEVAVTKASSVPAICRYGFTLAGDVGEFALVQPERKLPLOPA"  
complement(11717..3516)  
/locus\_tag="P2641"  
complement(11717..3516)  
/locus\_tag="P2641"  
/note="similar to SP:P37911, GB:X07469, SP:P08149,  
GB:M32091, GB:X07468, GB:X07470, GB:X59632, PID:150279,  
PID:44911, PID:44913, PID:44915, and PID:509155;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=1  
/product="iron-sulfur cluster-binding protein"  
/protein\_id="AA068249.1"  
/db\_xref="GI:24984219"  
/translacion="MKYFKLVLVLVDYSNNRSTICEVLNQSRHVIELSPGKTEPA  
SQEILALPMAASGIPVPSRCRAGCGCKXNVSGQDQKQADPTPSPPLAADM  
LILCSHACSDRLIEIPMSIDAPALFEOAVGFRASADIVELVLOPAPLEVRADQ  
LYRFLQNDGRCISILNLPARBOGQFHRKVGGLFIEGLIPTQAAATVLELP  
LGACTWDDDRPLILFATGAGVAGIFLLTLAAGDADVTWVGSGSPADFDREPL  
DVSRVHPHPMPVLSAORIOVALSOPHREMGEOVACGNATLMSQARQCLAG  
VPEPRFAEAPVASCALITTOASSANPILPELEKGPXYSLDGMIAAEQVRAYATA  
SQLDGATTAQALMAHATTOAKGASHWHEPTYIRFDQDVTYRPGQIDQRLVRLTD  
IIVVDVGPWMDYEGDYDTFVGQALHHAQVALHSEVDETRQAVRGILTGELYN  
FAERSAQKQKQERLNAGRIADFPVHLGODLAEVIEVPSVWVLEIOLCHPTE  
PVGAFFEDILANASRGIPGOLDYASALBPVTG"  
3517..5007  
/locus\_tag="P2642"  
3517..5007  
/locus\_tag="P2642"  
/note="Identified by match to PFM protein family HMM  
PF00155"  
/codon\_start=1  
/transl\_table=1  
/product="transcriptional regulator, GntR family"  
/protein\_id="AA068250.1"

/db\_xref="GI:24984220"  
/translation="MYQKRWQPLRLDQAQSYQIYEGLYAAISRGELRPTLPQ  
TRMAQLLVNREKTVILAEBAETKGLQSVQRGTPVSQLAPGTAAAGVRFAL  
ALPEBPAAVYFKANEQAVAOQGPALFPDNCORHLLPAVLRHYRANLRSFTS  
NSVRHSGESCRSLRSALADMLRNHNSLVGAHEHLCITGVQMSLSLVAYLKSGD  
TVVERLYSPAMEIFROLGARLYVELDHGGCVOLDALCROHKRMVYIPHPQF  
TVYSLAGRQOLLELRQAFQVVEEDYDEYHAGRPYPLASDROORHNYIGSL  
SKETSPRCSSYIVAEVYTVTLERPMILTGCPAQQMOCLADIINGELKKHRRV  
NREYRRRETLGCLDPAFGEQISVQAPREGGALAVYRPADENVQYLDKALERLVV  
RSGRQSPFGHAENLRIGFASLDDEIRQATQRLDAKAKIGKAR"  
5229. .6881  
/locus\_tag="PP2643"  
5229. .6881  
/locus\_tag="PP2643"  
/note="similar to GP:9947573; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="methyl-accepting chemotaxis transducer"  
/protein\_id="AA068251.1"  
/db\_xref="GI:24984221"  
/translation="MYPTSTARMLANLKRTGMFVYLSFSLTLFSTASAMWALG  
SDQITLEDQTAHQSDRLNNALIMAIRSANVSGFIHQGDHSDASGRNALSTEIN  
NKSQALVDFEVENAREPRLGLATELQATFAAYAAVAGQKATQRLDQYFKVNSD  
AGNAGRLQTLRQQLVTLISERGQOIMLESRRLARAOQLSLCLGVTVLWCMAR  
IAQVLPILREAGGHFRRIASGDLSPVQCGNNEIGQLFHLQMOOSORDLQGIN  
NCARQDAATALNVAIESANNLRQCGQELBOAATAVEMTAVEBARNAITSCQ  
TSESNQAAOSRROVSENIIDGTEAMREIQTSSAHLOQVGVROIGVLEIRVSE  
QNTLLALNATREAPARGAGRGAVADPRTLAFTQOSTEIDOMISVQAGTEAA  
VASQASTRAQSTLDVTLASGVLEGTSAIGTNERNLVLAABEQDAQYAREVDR  
NLNIRELSNHSAGAQOTSSEASKALSGVGMTALVGFKV"  
7131. .7277  
/locus\_tag="PP2644"  
7131. .7277  
/locus\_tag="PP2644"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AA068252.1"  
/db\_xref="GI:24984222"  
/translation="MLRDLPPVLAGMRQABPHPIRFTSPACVFCAAAGPRLAAS  
ARLP"  
7575. .10337  
/gene="mgcB"  
/locus\_tag="PP2645"  
7575. .10337  
/gene="mgcB"  
/locus\_tag="PP2645"  
/note="similar to GB:X60459, SP:P17181, PID:1247459,  
PID:1247461, PID:1567383, PID:1567385, PID:306914, and  
PID:32672; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="magnesium-translocating P-type ATPase"  
/protein\_id="AA068253.1"  
/db\_xref="GI:24984223"  
/translation="MKRLCVPPQROEHSNTVKYRNTSSAKARETSLSTRAAREAR  
NGIAVTLANDISEOGLTEHEAKRLVRGANOVADHPQPHLVQLLALNLPFVYL  
LTLAGISFVTDVYLPVSGADADADLTXYIIMTVASLSLRFQERSNKAALAK  
AMRTATVLRQIQAPRLREVPMDLVAQDIYQLSAGDMIPADILILERADLIFIS  
QAVLTGEALPVEKYDTLGNVAKSAEHAQODNLEFNICFGNTNVGAAVAVV  
ANGRRTYFGSLKAIGRSQTAQDFRGVSSVLLIRFLVWVVPVFMINGVKKDMA  
DAFLALAVAVGLTPMLPMIVSANAKAVAMARKVYVKLNLINLGSMDVLCCT  
KTGTLTQDRILILEHHVGPQGDYKRLILELAMSHHQSTRIILMOAYLHPGQHPQ  
QABYAYAKVDELDFDTRRRLLSVVNSALGDHLVSKGAEHMLAIAIHVGDDCYVA  
LDECRQQLMARVADANQDGFVLVATQIIPADSKAGVATHEDERLDLOGLLFTLD  
PKEATGPAIAALRDMGVOKVLTGDNPVATSVKREAGVLAGQPLDQDLEGMDIT  
LKQVEERTVFAKLTPLQKSRVLYKALQANGHTVFGSDINDAALRADVAVISDG  
TDIAKESADIIILERSIMYLEGVLKRGTEFGNIMKYLCTSSNFVGVYLVASR  
IPRLPMLATHILQNTMVDQSLSLPMQMDKEPLSKPKMDARNGIFPMIGTSS  
IPDITTFALMRYFRANSVQMQLRSGSHFIBGLLSQTLVYMLAKTRVYPPFOSTAL  
PVVLATGLVWALGIYIPFSPVGAMGVLPFWEIFPMLVATLLGICVVAQAKTLYIR

RFQGMF"  
10341. .10735  
/locus\_tag="PP2646"  
10341. .10735  
/locus\_tag="PP2646"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AA068254.1"  
/db\_xref="GI:24984224"  
/translation="MQGEQTRVLICAGRNVAUTRRQALDDVQRFQPIRVILHGS  
QFLGGEISWAREHAGDLYVPPNWLHGRQERLRLNLMINDSRDLIALFGDDT  
EELLARARCAELQVYVYAROPCEPDDCS"  
complement (10843, .12069)  
/locus\_tag="PP2647"  
/note="similar to GP:5738822; identified by sequence  
similarity; putative"  
/codon\_start=1  
Query Match 5.7%; Score 98.4; DB 1; Length 302101;  
Best Local Similarity 45.3%; Pred. No. 2.6e-08;  
Matches 718; Conservative 0; Mismatches 821; Indels 47; Gaps 8;  
125 TTAACACGCGCGGCAACCAAGAAACTATGCGCACTCGCCCTGACCCAGATCACTGCC 184  
DB TGACCAATGGCGTGGCGGTGACAGGGCCAGGCTACAGCCCGCTGACCTGCTCAATGTCA 44035  
QY 185 ACAACGTTGGTCACTTGCACCTGCTCGGCG-----CGCGGGATGGAGCGGGGG 225  
DB 44036 ATAAACGTCAGAGACCTCGCGCGGCTGGCGGCTTCTCTTCGCGGGAGAACACAGCGG 44095  
QY 236 CCGTACAGTCAACCGCATGATCATGATGATGATGATGATGATGATGATGATGATGATGATG 295  
DB 44096 GCGACAGCGCCACCGCGCTGATCAAGAGCGGGTGAATGATCAACCGCGCGCTGATCGG 44155  
QY 296 TGATCAGCGCGCTGATGCGCAACAGCGCATGATGATGATGATGATGATGATGATGATG 355  
DB 44156 GGGTGTTCGCGTGGATGCGCGCAACCGCAAGAACTGTGCAATGATGACGCTTGC 44215  
QY 356 CCGCGTGCACGCTAAACGCGCAAGCGCAAGCGCGCGCGCTTACCGCA 415  
DB 44216 ---CGATGACATCGCGCCCTGCTGCGACGTAATCAACCGCGCGCTGATCGCA 44272  
QY 416 CGACCTCTATTGAGCTCATGGGACCAACCATGATGCGCTGATGATGATGATGATGATG 475  
DB 44273 ACTGTGTTCTTTCGCGACCGCTGAGCGCAAGCTGTGGCGCTGAAACAAGACCGGCA 44332  
QY 476 AGCTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 535  
DB 44333 AGGTGCTGAGACAGAGAGGTGCGGACCAACAAGAGGCTACTCC--ATACGCGCG 44389  
QY 536 GAGCGATGTCGCCAATGAGGTGATGCTCGCGGTTCCACTTCCATATTCGCCCTATG 595  
DB 44390 CGCGATGATGCTCAATGGGCTAAGGCAAGTATCACAGCGGTGCGCGGAGATTCGGCGT 44449  
QY 596 GATCTTATATCTCGGGGACGATTCGCGGACGCGGTGAGAGCTGTGCGCAACCATTTA 655  
DB 44450 TGGGCAAGATTCAGGCGCTTACAACCCGAGAACGCGCAACTGTGTGATGCGGCCA 44509  
QY 656 TCCCGACGCGGGGGAAGAGGTGACAGACTTGCGGCAATGATTTGAGGCGCGCTGA 715  
DB 44510 TCGAAGGACACATGGGCTAAGTGTACAG--GATGCAAGCGGATTCGAGAACGATATTCC 44568  
QY 716 TGACCGCGCTTGGGGTCAATCACTATATTCGCGAGAACCTTTGTCTTCTAATGCT 775  
DB 44569 GCGCGTAGCGGGGCAAGCAACTGCGCGGCACTGTGAGAACCGCGCGCGCGCGCG 44628  
QY 776 CGACCGCGTGGCGCGCGGCGCGGCAACCCAGCGCGGCAAC-----GCGCGGCGGCA 827  
DB 44629 TGGCTGGGGGTTACTACGACCTCGAACCACCACTGATCTGTGTGGTACCGGTAAACCG 44688

QY 828 GCTGATGCAACAACACCCGTTTGCGTGCCTCCGACACGGGCGAGATTTGCTGCG 887  
 DB 44689 GCGCGGTGAACCTGCAACCTGCGCCCGGTGACAACTGTAATCTCTCTCAACGCTTGCA 44748  
 QY 888 T--CACCAGACCTCGCCCGGAGCAACTGGGACCAAGAACAGTTTGA-----G 936  
 DB 44749 CTGAACCCGAGCAGCGGACCATCAATGAGCATTTCCAGACGCGCATGACGGCTGG 44808  
 QY 937 ATGATGTCGCAACGTCATGATGCAACCCCGCCGAGATGAGAGGCTTGCAGCGCATC 996  
 DB 44809 GACTTCGACGCGCTCAAGAGACTATCTCTTCAACTCAAGAGCGCGGACAGAGAGTC 44868  
 QY 997 AACCCCAATGCGGCGAGCGGCGGAGCGCGCTGTGCTGAACGAGTGCCTTTCGAAGACCGGC 1056  
 DB 44869 AAGGCTCGCGCAGCGGAGACCGGACAGGTTTCTTCAACGCTGCAACCCACCAACGCGC 44928  
 QY 1057 ACGATGTCGTCGTTGATGCGCGGCTCGGGGGAATTCCTGTCGCGGCGGTATACCAATAC 1116  
 DB 44929 AAGTTCATCCGCGGCTTCCCTTCTGTCGACAAATCACTTGGGCTCACTGCGCTGACACAG 44988  
 QY 1117 ACCAATATGATCGCTCGATGACAGACCGGCTTGTGAACGAGTGAACGAGATGCGGTG 1176  
 DB 44989 GACGCGCGCGCGATCTCAACAGACGCGCGCGGCGGACACCGGACGAGGCGCAAG 45048  
 QY 1177 CTGAAGAGCTGGAACGTTGATATGACGTCTGCGCCGACCTTCTCTGGGTGGCGCGACTGG 1236  
 DB 45049 GGCAGCTCGGTG-----TTGCTGCGCGCGGCTTCTCTCGGCGCGCAAGAACTGG 45096  
 QY 1237 TCGTCAGCCGACATGAAACCCGACACCGGCACTTACTTCTTTCGCGCTGAAACAAATGCTGC 1296  
 DB 45097 ATGCGCATGCGCTCAACAAAGAGACAGAGGCTGTTCTACGTCGCGCTCAACAGAGTGGGCGC 45156  
 QY 1297 TACCATATATGCGCCGTTGATCAAGATTTAGCGCGCTGACGTCTATTAACCAAGCGCGC 1356  
 DB 45157 ATGACATCTGGAACGAAGGATCGCTATTAAGAAAGTCCGCGGTTCCTCGGTGCGCGC 45216  
 QY 1357 ACCGCAAACTCGCGCGCGGCTTGAATAATGCGCGCATTCACGCGCATGATATCAGC 1416  
 DB 45217 TTCAACATCAAGCCGCTCAATGAAGACTAATCGGCTGCTGCGCGCATCGACCGGTC 45276  
 QY 1417 ACCGCGCGCACCTTGTGTCGCGCGGCGCGCTGCGCGCAACTACCTGCCGTTTGTG 1476  
 DB 45277 AGCGGCAAGAGAGTGTGCGCGCAACAAGACTATCGCGCTGCGCGCTGCTGACAC 45336  
 QY 1477 ACCGCAAGCGGTGTGCTTCAACGCGGCGGACGACGCGCATTTCCGTGCGCTCAAGCAG 1536  
 DB 45337 ACCAAGGCAACCTGTGTGTCACGCGGACCGCAAGGGCTTCTTGAAGGCAATTCACGCC 45396  
 QY 1537 GAAACCGGCGAGACTTGTGTCAGAGCGGCTTTCGTCGACGCTGCGGACGAGGCGGATC 1596  
 DB 45397 AAGACCGGCGCAAGAGTGTGGAATTCAGACGCGGCTCGGCGGTGCTGCGCGGTC 45456  
 QY 1597 AGCTACAGTTGACGCGCTGCAATATATGCGCATTCGTCGCGCGGTCTGACCTATGCG 1656  
 DB 45457 ACCGGAAGAAAGACGCGCGGACATACGTTTCGCTGAGTCTCCGCTGCGCGCGCGGTG 45516  
 QY 1657 ACGCAATTGAACGCGCGCTGCGCA 1682  
 DB 45517 CCGCTGTGCGCGCGGCGGAAGTGGCAA 45542

RESULT 14  
 AY316747\_0/c  
 WPCOMMENT

Sequence split into 4 fragments LOCUS AY316747 Accession AY316747

Fragment Name

Begin End

AY316747\_0 1 110000

AY316747\_1 100001 210000

AY316747\_2 200001 310000

AY316747\_3 300001 357655

LOCUS AY316747 357655 bp DNA linear BCT 02-JAN-2004  
 DEFINITION Rhizobium sp. NGR234 megaplasmid 2 contig 1, complete sequence.

# ACCESSION VERSION KEYWORDS SOURCE ORGANISM

AY316747 GI:36958823  
 AY316747.1  
 Rhizobium sp. NGR234  
 Rhizobium sp. NGR234  
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.  
 1 (bases 1 to 357655)  
 Broughton, W.J., Perret, X., Staehelin, C., Schmitz, R.A., Raasch, C.,  
 Liesegang, H., Gottschalk, G. and Stett, W.R.  
 Comparative DNA analysis of two large contigs of the Rhizobium sp.  
 NGR234 megaplasmid 2  
 Unpublished  
 2 (bases 1 to 357655)  
 Broughton, W.J., Perret, X., Staehelin, C., Schmitz, R.A., Raasch, C.,  
 Liesegang, H., Gottschalk, G. and Stett, W.R.  
 Submitted (06-JUN-2003) Institut fuer Mikrobiologie und Genetik und  
 Laboratorium fuer Genomanalyse der Universitaet Goettingen,  
 Griesbachtstr. 8, Goettingen 37077, Germany  
 Location/Qualifiers

## FEATURES

source

## gene

/organism="Rhizobium sp. NGR234"  
 /mol\_type="genomic DNA"  
 /strain="NGR234"  
 /db\_xref="taxon:394"  
 /clone="contig 1"  
 /plasmid="megaplasmid 2"  
 complement(41..1449)  
 /locus\_tag="RNGR00476"  
 complement(41..1449)  
 /locus\_tag="RNGR00476"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Putative sensor histidine kinase protein"  
 /protein\_id="AA087601.1"  
 /db\_xref="GI:36959176"  
 /translation="MTKISDMLDPADNPENSTAKLTIAVLMKROVQASGREGVAY  
 EFOFPAALLERVRQRTDLSSTLKILNETNAPLAENARATERRARNLADLEAEEG  
 FALPFAISILMYNRSRPFCDLEDRATIRPGLSPADYIEACSRSPRLAPGGGADN  
 NATKQRRRTSVFQVALAGRWQVDSQPTSDGTVLVQTDVEILIMERSERML  
 DDQALIHATLEHINQICIFDAQRILVGMNERLRLDLIVTALRTGDTLQRM  
 QRDGLHSITPELFSMAVOQTPPLFLIRTSGLMVLVFAEETPDGTFWFSFD  
 VTRSELIOAMORANASLEAVAAETEDLRALADAEBSNTRAFVAAASHDLOPL  
 SAAKLFAVASDETERAGETLDPAKHALASVERALIGALLDISLDAVAVEISPVRL  
 APLFQGLTDERAPMAERKGLDIALIPCALTYFS"  
 complement(1446..2627)  
 /locus\_tag="RNGR00477"  
 complement(1446..2627)  
 /locus\_tag="RNGR00477"  
 /locus\_tag="RNGR00477"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Hypothetical protein"  
 /protein\_id="AA087249.1"  
 /db\_xref="GI:36958824"  
 /translation="WRIRFGSGVVRDAIVPSASTAANDPNAETLSRLQIGPSSLIM  
 LFIISADLERLIPAAARVFDVPTGCTTAGEISGAGYDEGTVALALPSEFRVDT  
 VTIPIMLDLPALAAVAVRKVAMARQYPELAEFALLVDGSKDEBLTSLAAG  
 LGAMPILGSGADGDRVARTFVLDGGRSNAALVIALVTRCPVRFVFDHMLPTERR  
 MYVTEADPRRVRVQINAPRAHEVAPLVGDPDOLSPFTFAHVAVAMKTRHHVRA  
 IROVNDNDLVFSAIDEGVYLTTERADNKGRLBELAGICMERPRPAILAEPICILR  
 RLEAEKQVIGRVSGILRRHGVGVSTYGEQFGPHVAVQMTGVALIPPGTVLPKQV  
 E"  
 complement(2701..3423)  
 /locus\_tag="RNGR00478"  
 complement(2701..3423)  
 /locus\_tag="RNGR00478"  
 /codon\_start=1  
 /transl\_table=11  
 /product="transcriptional regulatory protein"  
 /protein\_id="AA087250.1"  
 /db\_xref="GI:36958825"  
 /translation="MIIIRNSFQTPRTISVPQPLFKPYELRSALIVDDHPLFCALAM

## gene

CDS  
 (2701..3423)  
 /locus\_tag="RNGR00478"  
 complement(2701..3423)  
 /locus\_tag="RNGR00478"  
 /codon\_start=1  
 /transl\_table=11  
 /product="transcriptional regulatory protein"  
 /protein\_id="AA087250.1"  
 /db\_xref="GI:36958825"  
 /translation="MIIIRNSFQTPRTISVPQPLFKPYELRSALIVDDHPLFCALAM

gene  
CDS  
/locus\_tag="RNGR00479"  
/complement (3514..5280)  
/product="Alcohol dehydrogenase (acceptor)"  
/protein\_id="AA087251.1"  
/db\_xref="GI:36958836"  
/translation="MORIMLTATLTSLAGLASATVTEEDIAKDAATVGVDTNMGGR  
GLQRYPLELTKTNKRLPAWGSUGSKQKQSOPIVTDGMYTTASISRIAT  
DIRTKELWQDFARLPGLIPCCDIVNRGATLYGNTIFGTDLARLVLRKGTGVV  
NKKIANYKEGYSYTAAPLIVNGLITGNSGEGFIVEGVAQAKGELVWRPIEG  
HWSYKQKDSMTGNTLNAIWPGLAKTKGGATLGSYDADTDLVFGNGNAPMNSH  
LRNAKQVEENKGDNLVYAGRLGINPENGKIKHPQTTPREGMDPGVNEVPTISKD  
GTRPATADNNGSPFVILNPDGKFFVAAYPPVKNTAKGIDETGRPIYAEHRPDPS  
SAADGKQGVSEVSPFSGKMMPMAYESKSLPYPSNEMGMDITMNPISYKKGAA  
YLQAGFTIKPLPEDIYISLKAIDPNSGELIKMETKAAPILMGYMATRGIVVGTPEG  
EPRALDETEELMSFQTSQGVVGPITWQDEQYVSVISGWAIVPLMGSEVAKV  
NINQSGMWTFRLPKOLALAK"  
/locus\_tag="RNGR10005"  
/db\_xref="GI:36958832"  
/locus\_tag="RNGR10005"  
/codon\_start=1  
/translation="Coenzyme PQQ synthesis protein A"  
/protein\_id="AA087252.1"  
/db\_xref="GI:36958832"  
/translation="MSWHAQKFDVSCAMEITRYADADSEPLV"  
5908..6819  
/locus\_tag="RNGR00480"  
5908..6819  
/locus\_tag="RNGR00480"  
/codon\_start=1  
/translation="Coenzyme PQQ synthesis protein B"  
/protein\_id="AA087253.1"  
/db\_xref="GI:36958828"  
/translation="MKXASDLRVVLVGAAGGLPQMGCGNCNCAARPPSSLRPQT  
QSSLAVIDEAMALFNASPDIRQOIQVNAALQPHLRHSPIKAVLTNGDIDHLA  
LVIREKQPTFVSTGAVAEVIADNPVGPDLVSKSVAIIEPFSPLPGLARLFA  
VPGVPLFLEDEGDELGEQTVGELRADGRVYIYICGGLSGSLVAREADAL  
PFGCTESDEDEMIAGVGCKTGRMGMPVAGSDGLKALGPLIKTIKTYHINNTNP  
IMQGPBECAEAHGFVEGFPDGMENVR"  
6816..7583  
/locus\_tag="RNGR00481"  
6816..7583  
/locus\_tag="RNGR00481"  
/codon\_start=1  
/translation="Coenzyme PQQ synthesis protein C"  
/protein\_id="AA087254.1"  
/db\_xref="GI:36958829"  
/translation="MTIADKQAFHARLISGDRYHDKHPFHWMLHAGRCCTWQVRA  
WVLRVYQTRIPKMDAFLSCDDPQMRAMRSIEDHGGHDEGGRRMRLAEV  
VGLDADVASTGKVPATRPVADYVAVFVQCPLEAVASSTLELPAPKHSRIAGL  
LEHAYFADDAALAYFRRLVLEAPQDVFEGYLAAYLDHADTEKQNAVALTPTETVLM  
SQDALYGAIVAPGLIPGMDREGEVTAALAAAEAF"  
7580..7873  
/locus\_tag="RNGR00482"  
7580..7873  
/locus\_tag="RNGR00482"  
/codon\_start=1  
/translation="Hypothetical protein (probable Coenzyme PQQ  
synthesis protein D)"  
/protein\_id="AA087255.1"  
/db\_xref="GI:36958830"

gene  
CDS  
/locus\_tag="RNGR00483"  
/complement (7870..8994)  
/product="Coenzyme PQQ synthesis protein E"  
/protein\_id="AA087256.1"  
/db\_xref="GI:36958831"  
/translation="MNSKPIVSGDPAARHARVPPPMAMLAELTHRCPLACYSNPID  
LTHADGELSTGEMGVFEQAKGLVHLHLSGSPAPARDAELTRAAAFGYTNLI  
TSVGLTEARIGELADGLDVLQSLGSPGADRSIGRGYDRMAAFAVQVAT  
ELTVNAVGRONMDIDAMIDALIGRARIRIATVQFGMARNAALPRTVOVE  
RATTVAPARKYGVDEIDVADYHNSYPKCMGQMGVGLNTPSGRVLPOHAE  
TIFSLTDVNRKSLIATITREBDGFNAYRDTWMPELCRSCEKKNVDFGRCQAMAL  
AGDSASDPVCRSPRLGGLTEPDRLLASALVVRG"  
9259..10290  
/locus\_tag="RNGR00484"  
9259..10290  
/locus\_tag="RNGR00484"  
/codon\_start=1  
/translation="Transcriptional regulatory protein"  
/protein\_id="AA087257.1"  
/db\_xref="GI:36958832"  
/translation="MPENSTAHASFQRCSSPDSMETLGAETPDPAEIPISQARFR  
WGIDLSACDSAVLITGFHOTDFOHIEPTADTTEYSIIVPRSGKMTGYPTAEAG  
OGELLVYRNPNVIRGSENVDELITWVSVDLIGOTPEAPFPGSLDIPOVSL  
ATPAGQLIGNLFTIMGIRNDGLPLOSPLAMAHVYDALDVYRIRVPHLSLILXK  
PLIAPRHVRALIEFQANISRLTMMVVAAGVSVRALETGRARDDTPPAATYLR  
LRLAARQODLDPDNOPMRKICLKMGFHFGRFSAVRYSTYGENSDYTRRRYSSTA  
PYVFSDAQR"  
10561..11538  
/locus\_tag="RNGR00485"  
10561..11538  
/locus\_tag="RNGR00485"  
/codon\_start=1  
/translation="Transcriptional regulatory protein"  
/protein\_id="AA087258.1"  
/db\_xref="GI:36958833"  
/translation="MALGDPAARNPDWTFPSGDPDRPLATLTSTNSVYVPRPSEI  
SYKAMANAIGVSGSTHFPBQILLFSEBAGAGQVLIVPRRGVDLRMGQAICTBSG  
Query Match 5.5%; Score 95.2; DB 1; Length 110000;  
Best Local Similarity 44.0%; Pred. No. 1.1e-07;  
Matches 709; Conservative 0; Mismatches 873; Indels 30; Gaps 6;  
QY 149 ACTATCGGCACTCGCCCTGACCCAGATCACTGCGGCAACGCTGTCAGTTGCACTGG 208  
DB 5135 ACTCGCCCTCGAAACGTGAACAAGACGAAAGCTGAAGAACTCTGCTGCTGGGCTT 5076  
QY 209 TCTGGGCCCGCGGATGAGGCGGCGGCGGCGGATCAAGCATCAAGCCGATGATCATGATGGC 268  
DB 5075 TCAAGCTGGGGGGAAGGAGGCGGAGCAAGAAAGCGAGCCGATCGTCAATATGGCG 5016  
QY 269 TGATTTATCTGGCAAAACCCCGGATGATGATCAAGCGGCGTGAATGCCCAACAGCGCATC 328  
DB 5015 TGATTTATTAACCGCCCTCAACGCGGCTTTAGCCATCGATGATCAAGCGGGAAGG 4956  
QY 329 TGATCTGGGAACACGCGCGCAACTGCGCGCGCTGCGCACGCTAAAGCGCCCAAGCGCAC 388  
DB 4955 AGCTTGCGCATTTGCAAGCCCGCTTGCCCGAGGATATCTGCGCTGTGGGA---CGTGA 4899  
QY 389 GCAAGCGCGCGGTGCGCTTTAAGCGCACAGAGCTCTATTTCAGCTCATGGACAACATC 448  
DB 4898 TCAACCGCGGTGCGCGCATTCACGCGCAATATCTATTTCGAGCAACCTGATCCCGTC 4839  
QY 449 TGATCGCGCTGATATGAGACGCGGCGCAGTCTGATTCGATGTCGAGTGGATCGGCG 508  
DB 4838 TGTGCGCCCTCAACCGCAAGCGGCGATGTGCTGTGGAACAAGATGCCCAATTACA 4779

QY 509 AAGACGGCTTGAACAGTAAACAGGAGGCGGATGTCGCAATGAGCGATGTCGCG 568  
 DB 4778 AGGAAGGCT---ACAGCTAACAGGCGCGCGCTTGATCGTGAACGGCGCTCATATCAACGG 4722  
 QY 569 GTTCCACCTGACCAATATTCGCGCTATGATGCTTTATCTGAGGAGGACGATTCGCGACCG 628  
 DB 4721 GCAATTCGGGTGGCGATTGCGTATGCTGCGGAGGATTCAGAGCGCGCGATGCGCAACCG 4662  
 QY 629 GTGAGAGCTGTGGCGCAACCACTTTATCCGAGCGGCGGAGAGAGGATGACGACTT 688  
 DB 4661 GCGAATCGTATGAGACCCGCTCGGTCATGAGAGGACGTCGGGATTTCAAGGGCAAGG 4602  
 QY 689 GGGGCAATGATTTGAGAGCGCGCTGGATGACCGG---CGTCTGGGATCAAGTACCTATG 745  
 DB 4601 ATAGCTCGATGACCGGAGCGCTGATGATGCGACCTGGCCCGCGACCTTTGGAAGACAGCG 4542  
 QY 746 ATCCCGTGAACGAACTTGTGTTCTATGAGCTGACCGCGCGTGGGCGCAGGCGTCCGAACCC 805  
 DB 4541 GCGGCGGCACTGCGCTCGCGGATCTTACAGCGCGATACCGATACGCTGCTTTCGGCA 4482  
 QY 806 AGCGCGGACGCGCGGCGGACGCTGATGACCAACACCGCTTTCGCGTCCGTCGCG 865  
 DB 4481 CCGGCAACCCCGCTCTTGGAGACGCACTCTGCGCAACCGCGCAAGCGGTGAGGGCA 4422  
 QY 866 ACAAGGCGGAGATTGCT-----GGCGTACCGAGACCTTGGCGCGCGCAACTGGGAC 919  
 DB 4421 ACAAGGCGGCAACCTCTACCGCGCGAGCGCTTCGCAATCAATCCGAGAAATGCGGAGA 4362  
 QY 920 AAGATGCAAGCTTGAAGATGATGCGCA-----CGTGAATGTCACACCT 967  
 DB 4361 TCAAGTGCACCTTCCAGACGACCCGAGGAGGAGCTGGGATTTGACGCGTGTGAATGAG 4302  
 QY 968 CGGCGGAGATGAGAGGCTGCGCGCATCAACCCCATGCGGCGAGCGCGCGCGCTG 1027  
 DB 4301 TCGTGCCTTCATCTCCAAAGAGGAGGACGAGAGCGCTTGGACCGCGCAACCGAAAGCT 4242  
 QY 1028 TCGTGAAGGCTGCGCTTGAAGAGCGGACGATGCTGCTTGAATGCGGCTCGGCG 1087  
 DB 4241 TCTTCAAGCTTCAACCGGAGAGGAAAGTTCTGCGCGCTTCTTTCGTTTAA 4182  
 QY 1088 AATTCCTGTGGCGCGCTGATCAACCACTCAACCAATGATGCTGATGACGAGACCG 1147  
 DB 4181 ACATCACTGCGGAGAGGATGACGAGACCGGTGCGCGCATCTATGCGGAGAAACC 4122  
 QY 1148 GCCTTGAACGATGAACGAGATGCGGTGCTGAAAGAGCTGACGTTGAATGACGCT 1207  
 DB 4121 GCGCCGCGCATCTGCTTTCGCGCGCGAGCGCGCAAGGGGCGAGCGTCTTTGCGT-- 4064  
 QY 1208 GCGCGACCTTCTGAGTGGCGGAGCTGCTGCTGACCGCACTGAACCGCGACCGCGCA 1267  
 DB 4063 -TCCCTCTCTCTCGGCGGCAAGACTGAGTGGCGCTACAGCGAAGAGAGGAGC 4005  
 QY 1268 TTTACTTCTTGGCGCTGGAACATGCTGCTGACATTTATGAGCGCTTGAATCAAGCTTTA 1327  
 DB 4004 TCTTTTACGTCCTCGAACGAAATGAGATGATCTGGAATGAGCGGATCTCTTACA 3945  
 QY 1328 GCGCGCTGACGCTTAAACACAGCGCGACCGCAAACTCGCGCGCGGCTTTGAATA 1387  
 DB 3944 AGAAGGGGGCGGCTACCTCGGGCGGCTTCAACATCAAGCGCGCTTTCGAAGCTTAA 3885  
 QY 1388 TGGGCGGATGACGCGATTTGAATACGACCGGGCGCACCTTGTGTGCGGAGCGCC 1447  
 DB 3884 TCGGCTCGCTGAAGGCGATGACCGCAACGCGCGAGATCAAGTGGGAATCAAGAAAG 3825  
 QY 1448 CTGCGGGAACCTACTCGCGCGCTTTTGTGCGCGGACGCGGTGTGTTCAACGCGGGA 1507  
 DB 3824 CCGCGCGCGCTGTGGGCGGTGTATGCGCACCTGCGCGGGGCGTGTCTTCTGCGAAGCG 3765  
 QY 1508 CCGACGCGATTTTCCGTCCTCAGCCGAGAAACCGGCGAAGCTTGTGGGAGCGCGCTC 1567  
 DB 3764 CCGAGGCGGATTTCAAGGCGCTGACGAGAGACCGGCGAAGAGCTCTGTCTTTCAGA 3705

QY 1568 TTGCAAGCTGCGACGCGGAGGAGGATGACGATGAGATTGAGACGCGGTGCAATATATCG 1627  
 DB 3704 CCGGCTCGGCGGATGATCGCGACCGGATCACTGAGAGCAGACGCGGAGCAATATGTGT 3645  
 QY 1628 CCATGGGTGCGGCGGTGATGACCTATGAGCAGCAATTGAAAGCGCGCTGCGCAGGCA 1687  
 DB 3644 CGGTATCTTCGGGTGGGCGCGCGGTCCCTTTTGGGCGGCGAAGTCCCAAGAGG 3585  
 QY 1688 TCGATTGACCTCGGTGCGTATGCGATCTATGCTTTTTCACCTGCGCGAGTA 1739  
 DB 3584 TCAATATCTCAACCAAGTGGCATGTTTGGACTTCCGCTTCCGAGAGA 3533

## RESULT 15

AF35798

2550 bp DNA linear BCT 14-MAR-2002

## LOCUS

Pseudomonas butanovora 1-butanol dehydrogenase BDH precursor, gene,

## DEFINITION

complete cds.

## ACCESSION

AF35798.2 GI:16579769

## VERSION

AF35798.2

## KEYWORDS

Pseudomonas butanovora

## ORGANISM

Pseudomonas butanovora

## REFERENCE

Bacteria; Proteobacteria; Betaproteobacteria; Rhodocyclales; Rhodocyclaceae; Thauer.

## AUTHORS

1 (bases 1 to 2550) Vangnai, A.S., Arp, D.J. and Sayavedra-Soto, L.A.

## TITLE

Two distinct alcohol dehydrogenases participate in butane metabolism by Pseudomonas butanovora

## JOURNAL

J. Bacteriol. 184 (7), 1916-1924 (2002)

## MEDLINE

21886383

## PUBMED

11890998

## REFERENCE

2 (bases 1 to 2550) Vangnai, A.S., Arp, D.J. and Sayavedra-Soto, L.A.

## AUTHORS

Direct Submission

## TITLE

Submitted (02-MAR-2001) Botany and Plant Pathology, Oregon State University, Corvallis Hall, Corvallis, OR 97331, USA

## JOURNAL

3 (bases 1 to 2550) Vangnai, A.S., Arp, D.J. and Sayavedra-Soto, L.A.

## REFERENCE

Direct Submission (01-NOV-2001) Botany and Plant Pathology, Oregon State University, Corvallis Hall, Corvallis, OR 97331, USA

## AUTHORS

Sequence update by submitter

## REMARK

On Nov 1, 2001 this sequence version replaced gi:13448822.

## COMMENT

Location/Qualifiers

## FEATURES

1. 2550

## source

/organism="Pseudomonas butanovora"

## CDS

/mol\_type="genomic DNA"

## CDS

/db\_xref="taxon:86174"

## CDS

369..2444

## CDS

/codon\_start=1

## CDS

/product="1-butanol dehydrogenase BDH precursor"

## CDS

/protein\_id="AAK27220.2"

## CDS

/db\_xref="GI:16579770"

## CDS

/translation="MLTTFPARKREESVPLRKIGORALLGLSCVLTSTTSFAAGGEMR

## CDS

THGYDDAGTIRYSLAQITPDNAKELIWMVSVDLSSRGVATPIVVDGVMYVIAFMV

## CDS

VHALDVSGRLMTYDPEPREKKNACCVNNGVAGVAHSGKVFGLDGLVADAR

## CDS

TGRVWRNTLIDDDKPYITTGAPRVKIGVINGGAEGRVGYITAPDPAASRG

## CDS

VTPGPDPSLPRDMSWEAAKTMDDPAQVIGSGRSPRGVGLDILYKAKPCTTSARA

## CDS

TPSPMSRKSPRAGDNLTYASTYALAPDPTGRTYMHQCPADNMWDTSTODILADI

## CDS

ELGGKRPVLIHAKNGFFVIDRTDKTISAQNFVFNATGIDENGRLENBEGAM

## CDS

PGLHSRPPAPASRTNWSYSPQTLAEFPQNTPLVIOEDKMSVNOAQPOAMA

## CDS

GIGNMLMLVNPQEPGRLIAMPVOQKEVWRKEHVS PMNGTLLVTAGNVVFG

## CDS

TADRLIAPFARDKELMSAPMGTVIAAPVYEVDEKQVSIAVGGVYGNFTRAS

## CDS

ERRPTGVTFALGKEMPAFTFVOLNMLVSGVDVYPDVAVSGTGLYTNVCVCHGV

## CDS

PCNDKGGGIPNLGTAETIAHLDQFVFKRPFMRGMPDPTTKLTLPQVEKIKAFILG

## CDS

TADVVRPK"

## ORIGIN

Query Match 5.4%; Score 94.8; DB 1; Length 2550;  
 Best Local Similarity 48.3%; Pred. No. 2e-07;  
 Matches 452; Conservative 0; Mismatches 442; Indels 42; Gaps 5;

QY	105	CCCGCCGGCTGTGAATGAGTTAACTAAGCGCGCAACCAAGAAAACTTATGCGCACTGCC	164
Db	479	CGTGGCCGTGGCGAATGGCCGACGCATGGTTTATGACACGCCCGGACGCGCTACAGCC	538
QY	165	CCTGACCCGATCACTGTCCGACCAACGTTGGTGTCAAGTTGCACTGGTCTGGGCCCCGGGAT	224
Db	539	CCTGGCCAGATCAAGCCGCGCAACATGCCAAGAACTCCGGGCTGGTCTGGTCTTACGACT	598
QY	225	GAGAGCCGGGG---GCCGTTACAGGTTCAGGCCGATAGATCATATGAGCGGATGTATCTAGC	281
Db	599	GGAGTGTGTGGCGGGCGTGCAGAACCCAGCCGAACTCTGGTCAATGGCGTGAATGTAAGTAC	658
QY	282	AAACCCCGGTATGTATGATTCAGGCGCGCTGGATTCGCGMAACAGCGCATCTGATCTGGGACA	341
Db	659	CGCGCCGTGGAGCGGTGTGTGATGGCTGTGACGTGGCGACGGCAAGCGGTTGTGGACTTA	718
QY	342	CCGCGCCCAACTGCGCCCGCGCTGTGCCACGCTAAACGCCCAAGGCCA---CCGACAGCGCG	398
Db	719	CGACCCGGAAGTTCCGCGCGAGAAAGGCGCAAGATATCTTTGGTCGACGTGTGAACCGCG	778
QY	359	CGTGGCCCTTTAAGCGCACAGAGCCTTATTTAGCTCATGGGACACAACATCTGATGCGACT	458
Db	779	CGTGGCCGTGATAGAGGGCAGAGTTTGTGTGGCGTCCGTTCGACGGCCGGCTGGTCCGAT	838
QY	459	GGATATGAGACGGGCGACAGTGTCTATTTGATGTGCAACGTGATCGGCGCGAAGCGCTT	518
Db	839	CGACCCCAAGACCCGCAAGCGGGTGTGGAGCGGAAACACGCTGATGACGACGACAAAGCC	898
QY	519	GACCAGTAAACACCAAGGGGCGCATTTGTGGCCCAATGGCCTCACTGTCGGGAGTTCCACCTG	578
Db	899	CTACACCACTACCGGTTGGCGCCCGCGTCAATCAAGGGCAAGGTCGTCACTGGCAACGCGCG	958
QY	579	CCAAATATTGCGCCCTATGATGTCTTATCTCGGGGACGA-----617	
Db	959	TGCGGAGTTCCGCGCGTGGCGCGCTACATACCGCCCTTACACCCGACAGCGCGAGAGGCC	1018
QY	618	-----TTCCGGGACGGGGTGAAGAGCTGTGGGCGCAACCACTTATTCG-----GCAGC	664
Db	1019	TGGCGTGTTCCTCGGCGCGGGTGTGATCCGTGCTGCCGTTGCAAGATGCTCTGATGAGAGGC	1078
QY	665	CGGGCGAAGAGGTTACGAGACTTTGGGCAATGATTTGAGCGCGCTGGATGACCGGCG	724
Db	1079	AGCGGCGCAAGACTCTGGGACCCGCGCGGCAAGTATGGGAAAGCGGGCGCGCGGACGG	1138
QY	725	TCTGGGGTCAAGTCAACCTATGATTCGCCGAGCAACCTTGTTGTTAT--GGCTCGACCGG	782
Db	1139	TGTGAACTCGATGGTTTATACCGCAAAAGCTGGATTCTGTTGTACATGGCAGCGGCAAC	1198
QY	783	CGTGGGCCAGCGCTCGGAACCCAGACCGCGGCAACCGCGGGCGGCAACGCTGTATGGCACCCA	842
Db	1199	CCCCCTGGCGGTGTGCAACCGCAACGTAAAGCCCGGCGGCGGCAACCTGTACACGCG	1258
QY	843	CACCCGCTTTGGGTTGGCGTCCCGACACGGGGGAGATTTGTTCGGCTCACAGACCTTGGC	902
Db	1259	GTTCATCTGTGGCGTGGCGCGCGGAGACACGGGCGCAATACGTTCTGACCTTACAGCAAGCC	1318
QY	903	GGCGCACAACTGGGACCAAGATGCACTTCCAGATGATGTGTGCGCAACGTGATGTGCA	962
Db	1319	GGCGGACAACTGGGACTTACCTCGACGAGGAACTGATCTCGGCGCACTGCAACTGGG	1378
QY	963	ACCTTCGGCCGAGATGAGAGGTTCTGGCGCATCAA	998
Db	1379	GGGCAAGCGCGCAAGTCACTCTTCATATGGCGCCA	1414





recombinant organisms can be used to convert suitable substrates to aldehydes, ketones or carboxylic acids, especially to convert L-sorbose or D-sorbitol to 2-keto-L-gulononic acid, which can be converted to L-ascorbic acid by standard procedures. The derivatives of ADH enzymes have desired substrate specificity, higher affinity to a substrate, lower affinity to an inhibitory compound, higher stability against temperature and/or pH and higher catalytic speed

Sequence 1740 BP; 340 A; 529 C; 545 G; 326 T; 0 U; 0 Other;

Query Match 100.0%; Score 1740; DB 2; Length 1740;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGACCCCAACAGCTGCTTGCACACAGCGCGCGCTGATTTGCTTACCGCGCCGCC 60
DB 1 ATGACCCCAACAGCTGCTTGCACACAGCGCGCGCTGATTTGCTTACCGCGCCGCC 60
QY 61 GCATTGGCGGAGTAACCCCGATTACCGATGAACTCTGCGCAACCCCGCGCTGTGA 120
DB 61 GCATTGGCGGAGTAACCCCGATTACCGATGAACTCTGCGCAACCCCGCGCTGTGA 120
QY 121 TGAATTACTACGCGCGCAACCAAGAAACTATCGCACTCGCGCTGACCGAGATCACT 180
DB 121 TGAATTACTACGCGCGCAACCAAGAAACTATCGCACTCGCGCTGACCGAGATCACT 180
QY 181 GCCGACAACGTGTGCTGCTGCACTGTGTGCGCGCGCGGATGAGCGCGCGCGCTA 240
DB 181 GCCGACAACGTGTGCTGCTGCACTGTGTGCGCGCGCGGATGAGCGCGCGCGCTA 240
QY 241 CAGGTACAGCGCGATGATCCATGATGAGCGGATGATCTGCGCAACCCCGGATGATC 300
DB 241 CAGGTACAGCGCGATGATCCATGATGAGCGGATGATCTGCGCAACCCCGGATGATC 300
QY 301 CAGGCGGTGATGACGCAACAGCGGATGATCTGCGCAACCCCGCGCACTGCGCGC 360
DB 301 CAGGCGGTGATGACGCAACAGCGGATGATCTGCGCAACCCCGCGCACTGCGCGC 360
QY 361 GTGCGCACGCTAAACCGCGCAAGCGGACCGCAACCGCGCGCGCTTACCGCGCAAC 420
DB 361 GTGCGCACGCTAAACCGCGCAAGCGGACCGCAACCGCGCGCGCTTACCGCGCAAC 420
QY 421 CTCTATTTCACTCATGAGCAACCATCTGATCGCGCTGATATGAGACGCGCAAGTC 480
DB 421 CTCTATTTCACTCATGAGCAACCATCTGATCGCGCTGATATGAGACGCGCAAGTC 480
QY 481 GTATTGATGTGCAAGCTGATGCGCGCAAGACGCTTGAACAGTAAACCAACGCGCGC 540
DB 481 GTATTGATGTGCAAGCTGATGCGCGCAAGACGCTTGAACAGTAAACCAACGCGCGC 540
QY 541 ATTGTGCGCAATGCGCTGATCGCGGCTTCACTGCGCAATTTGCGCTATGGAAGC 600
DB 541 ATTGTGCGCAATGCGCTGATCGCGGCTTCACTGCGCAATTTGCGCTATGGAAGC 600
QY 601 TTTATCTCGGGGACGATTCGCGGACGCGGATGAGACCTGTGCGCAACCACTTTATCCG 660
DB 601 TTTATCTCGGGGACGATTCGCGGACGCGGATGAGACCTGTGCGCAACCACTTTATCCG 660
QY 661 CAGCGCGGCAAGAGGATGACGACTTGGGCAATGATTTGAGGCGCGCTGATGAC 720
DB 661 CAGCGCGGCAAGAGGATGACGACTTGGGCAATGATTTGAGGCGCGCTGATGAC 720
QY 721 GCGGTCTGGGGTCAATCACTTATGATCCGTGACGAACCTTGTTCTATGCTGAC 780
DB 721 GCGGTCTGGGGTCAATCACTTATGATCCGTGACGAACCTTGTTCTATGCTGAC 780
QY 781 GCGGTGGGCGGCGGACGCGGCAACCGCGGCGCGCGCGCGCGCGCTGATGAC 840
DB 781 GCGGTGGGCGGCGGACGCGGCAACCGCGGCGCGCGCGCGCGCGCTGATGAC 840
QY 841 AACACCGCTTTGCGGTGCTGCCGACACGCGGAGATTTGCTGCGGTACACGACCTG 900
DB 841 AACACCGCTTTGCGGTGCTGCCGACACGCGGAGATTTGCTGCGGTACACGACCTG 900
```

```
QY 901 CCGCGCAACAATGGGACCAAGATGACCGCTTGAGATGATGTGCGCAACGTGATGTG 960
DB 901 CCGCGCAACAATGGGACCAAGATGACCGCTTGAGATGATGTGCGCAACGTGATGTG 960
QY 961 CAACCTTCGCGCGGATGAGGAGGTCTGCGCGCATCAACCCCAATGCGGGAAGCGGAG 1020
DB 961 CAACCTTCGCGCGGATGAGGAGGTCTGCGCGCATCAACCCCAATGCGGGAAGCGGAG 1020
QY 1021 CGCGGTGTGCTGACGAGGTGCGCTTGCAGAACCGGCAAGATGTGTGATGCGGCC 1080
DB 1021 CGCGGTGTGCTGACGAGGTGCGCTTGCAGAACCGGCAAGATGTGTGATGCGGCC 1080
QY 1081 TCGGCGCAATTTCTGTGCGCGCGGTGATCAACACTACCAATATGATCGCTGATGAC 1140
DB 1081 TCGGCGCAATTTCTGTGCGCGCGGTGATCAACACTACCAATATGATCGCTGATGAC 1140
QY 1141 GAGACCGGCGCTTGTGACGATGAAAGAGATGCGGTGTGAAAGAGCTGACGTTGATAT 1200
DB 1141 GAGACCGGCGCTTGTGACGATGAAAGAGATGCGGTGTGAAAGAGCTGACGTTGATAT 1200
QY 1201 GACGTGTGCGGACCTTCTGCGGTGCGCGGACGCTGCTGACGCGCACTGAACCGGAC 1260
DB 1201 GACGTGTGCGGACCTTCTGCGGTGCGCGGACGCTGCTGACGCGCACTGAACCGGAC 1260
QY 1261 ACCGCAATTTACTTCTGCGCGCTGAAACAATGCGCTACGATATTTAGCCGTTGATCA 1320
DB 1261 ACCGCAATTTACTTCTGCGCGCTGAAACAATGCGCTACGATATTTAGCCGTTGATCA 1320
QY 1321 GAGTTTACGCGCTGCAAGTCTATTAACCAAGCGCAACCGCAAACTGCGCGCGCTTT 1380
DB 1321 GAGTTTACGCGCTGCAAGTCTATTAACCAAGCGCAACCGCAAACTGCGCGCGCTTT 1380
QY 1381 GAAATATGCGCGCGCAACGCAATGATATGAGCAACCGGCGGCACTGTGCTGCGG 1440
DB 1381 GAAATATGCGCGCGCAACGCAATGATATGAGCAACCGGCGGCACTGTGCTGCGG 1440
QY 1441 GAGCGCGCTGCGGGAACCTACGCGCTTTGTGCAACGCGAGCGGTGTGATTTCAAC 1500
DB 1441 GAGCGCGCTGCGGGAACCTACGCGCTTTGTGCAACGCGAGCGGTGTGATTTCAAC 1500
QY 1501 GCGCGGACCGACGCTATTTTCGCTGCTGACGCAAGAAACCGGCGAGCTTTGTGCG 1560
DB 1501 GCGCGGACCGACGCTATTTTCGCTGCTGACGCAAGAAACCGGCGAGCTTTGTGCG 1560
QY 1561 GCCGCTCTGCGACGCTGCGGACGCGGCAAGGATCACTACGATTTGACGCGCGTCA 1620
DB 1561 GCCGCTCTGCGACGCTGCGGACGCGGCAAGGATCACTACGATTTGACGCGCGTCA 1620
QY 1621 TATATGCGCATGCGGTGCGGCGGTCTGACCTATGCGCAACGCAATTGAACGCGCGT 1680
DB 1621 TATATGCGCATGCGGTGCGGCGGTCTGACCTATGCGCAACGCAATTGAACGCGCGT 1680
QY 1681 GAGGCAATGATGATGACCTGCGGTGCTATGCGATGCTTTGCACTGCGCGCAGTAA 1740
DB 1681 GAGGCAATGATGATGACCTGCGGTGCTATGCGATGCTTTGCACTGCGCGCAGTAA 1740
```

## RESULT 2

ADN10955 standard; DNA; 1740 BP.

ADN10955;

01-JUL-2004 (first entry)

Glucobacter oxydans Enzyme B coding sequence.

Enzyme B; ascorbic acid; vitamin C; L-gulono-1,4-lactone; L-gulononic acid; L-galactono-1,4-lactone; L-galactonic acid; gene; ds.

Glucobacter oxydans.

FH Key Location/Qualifiers  
 FT CDS 1..1740  
 FT /\*tag= a  
 FT /product= "Enzyme B"  
 PN WO2004029267-A1.  
 XX 08-APR-2004.  
 XX 22-SEP-2003; 2003MO-EP010489.  
 XX 27-SEP-2002; 2002EP-00021602.  
 XX (STAM ) DSM IP ASSETS BV.  
 PA Hoshino T, Shinjoh M;  
 PI WPI; 2004-329889/30.  
 DR P-PDB; ADNI0956.  
 DR  
 XX  
 PT Producing L-ascorbic acid using enzyme B of Gluconobacter oxydans, from  
 PT substrates L-gulose, L-galactose, L-idose, and L-talose.  
 XX  
 PS Disclosure; SEQ ID NO 1; 24bp; English.  
 XX  
 CC The present sequence is the coding sequence for Enzyme B of Gluconobacter  
 CC oxydans strain DSM 4025. Enzyme B has a molecular weight of about 60,000  
 CC Da by SDS-PAGE, substrate specificity for primary and secondary alcohols  
 CC and aldehydes, is stable in the pH range 6-9 with optimal activity at  
 CC about pH 8.0, and is inhibited by Cu<sup>2+</sup>, Zn<sup>2+</sup>, Mn<sup>2+</sup>, Fe<sup>2+</sup> and Fe<sup>3+</sup>. The  
 CC present invention provides the use of this enzyme in a process for  
 CC producing L-ascorbic acid from L-gulose, L-galactose, L-idose or L-  
 CC talose, or from L-gulono-1,4-lactone, L-gulonic acid, L-galactono-1,4-  
 CC lactone, L-galactonic acid, L-idono-1,4-lactone, L-idonic acid, L-talono-  
 CC 1,4-lactone and L-talonic acid. Enzyme B is also used in a process for  
 CC the production of L-gulono-1,4-lactone or L-gulonic acid from L-gulose,  
 CC and L-galactono-1,4-lactone or L-galactonic acid from L-galactose. The  
 CC processes involve contacting the enzyme with the respective substrate and  
 CC isolating the product from the reaction mixture. The process is conducted  
 CC for 1-20 hours at pH 1-9 (preferably pH 2-8) and 13-45 (preferably 18-42)  
 CC degrees C. Production of L-gulono-1,4-lactone/L-gulonic acid from L-  
 CC gulose, vitamin C from L-gulono-1,4-lactone/L-gulonic acid, L-galactono-  
 CC 1,4-lactone/L-galactonic acid from L-galactose, and vitamin C from L-  
 CC galactono-1,4-lactone/L-galactonic acid by recombinant *Escherichia coli*  
 CC JM109 carrying the Enzyme B gene is described in examples from the  
 CC invention.  
 CC  
 XX Sequence 1740 BP; 340 A; 529 C; 545 G; 326 T; 0 U; 0 Other;

Query Match 100.0%; Score 1740; DB 12; Length 1740;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACCCCAACAGCTGCTTGCACACGCGCGCTGCTATTGGTTACCGCGCCGCC 60  
 Db 1 ATGAACCCCAACAGCTGCTTGCACACGCGCGCTGCTATTGGTTACCGCGCCGCC 60  
 QY 61 GCATTGCGCGAGTAACCCCGATTACCGATGAACTCTGCGAACC CGCGCTGGTGA 120  
 Db 61 GCATTGCGCGAGTAACCCCGATTACCGATGAACTCTGCGAACC CGCGCTGGTGA 120  
 QY 121 TGAATTACCTAGCGCGCAACCAAGAAACTATCGCACTCGCCCTGACCCAGATCACT 180  
 Db 121 TGAATTACCTAGCGCGCAACCAAGAAACTATCGCACTCGCCCTGACCCAGATCACT 180  
 QY 181 GCCGACAACGTTGGTCACTGATCTGGGCGCGCGAGTGAAGCGGCGGCGCTA 240  
 Db 181 GCCGACAACGTTGGTCACTGATCTGGGCGCGCGAGTGAAGCGGCGGCGCTA 240  
 QY 241 CAGGTCAACGCGATGATCCATGATGGCGTGTATCTGGCAAAACCCCGGTATGTATC 300  
 Db 241 CAGGTCAACGCGATGATCCATGATGGCGTGTATCTGGCAAAACCCCGGTATGTATC 300

QY 301 CAGCGCTGATGCGCAAAACAGCGATGTATCTGGGAACACCGCGCGCAACTGCGGCC 360  
 Db 301 CAGCGCTGATGCGCAAAACAGCGATGTATCTGGGAACACCGCGCGCAACTGCGGCC 360  
 QY 361 GTCCGCAACGCTAAACGCCCAAGCGGACCGCAAGCGCGCGCTGCGCTTACCGCAGAGC 420  
 Db 361 GTCCGCAACGCTAAACGCCCAAGCGGACCGCAAGCGCGCGCTGCGCTTACCGCAGAGC 420  
 QY 421 CTCTATTTCACTGATGGAACCAACATCTGATCGCGCTGATATGAGACCGGCGAGTC 480  
 Db 421 CTCTATTTCACTGATGGAACCAACATCTGATCGCGCTGATATGAGACCGGCGAGTC 480  
 QY 481 GTATTGATGTCGAACCGTGAATCGGGGGAACACGGCTTGACCAATACACACGGGCGCG 540  
 Db 481 GTATTGATGTCGAACCGTGAATCGGGGGAACACGGCTTGACCAATACACACGGGCGCG 540  
 QY 541 ATTGTCGCAATGCGATCATGTCGCGGGTTCCACTGCGCAATATTCGCCCTATGATGC 600  
 Db 541 ATTGTCGCAATGCGATCATGTCGCGGGTTCCACTGCGCAATATTCGCCCTATGATGC 600  
 QY 601 TTTATCTGGGGCACAGATTCGCGACGCGGTGAGAGCTGTGGCCCAACACTTTATCCGC 660  
 Db 601 TTTATCTGGGGCACAGATTCGCGACGCGGTGAGAGCTGTGGCCCAACACTTTATCCGC 660  
 QY 661 CAGCGCGCGAAGAGGGTGAAGAGCTTGGGGCAATGATTGAGAGCGCGCTGATGAC 720  
 Db 661 CAGCGCGCGAAGAGGGTGAAGAGCTTGGGGCAATGATTGAGAGCGCGCTGATGAC 720  
 QY 721 GCGGCTGGGGTCAAGATCACTATGATCCGTGACGAACCTTGTGTTATGCTGACAC 780  
 Db 721 GCGGCTGGGGTCAAGATCACTATGATCCGTGACGAACCTTGTGTTATGCTGACAC 780  
 QY 781 GCGGCGGGCCAGGGTCCGAACCCAGCGGCGGCGCGCGCGCGCGCTGATGACAC 840  
 Db 781 GCGGCGGGCCAGGGTCCGAACCCAGCGGCGGCGCGCGCGCGCTGATGACAC 840  
 QY 841 AACACCGCTTTGGCGTCCGTCGCCGACACGCGGCGAGATTGCTGCGCTCACAGACCTG 900  
 Db 841 AACACCGCTTTGGCGTCCGTCGCCGACACGCGGCGAGATTGCTGCGCTCACAGACCTG 900  
 QY 901 CCGCGCGCAACTGGGACCAAGATGACGTTCCAGATGATGTGCGCAACGTCGATGTG 960  
 Db 901 CCGCGCGCAACTGGGACCAAGATGACGTTCCAGATGATGTGCGCAACGTCGATGTG 960  
 QY 961 CAACCTCGCGCGAGATGAGAGGTCCTGCGCGCATCAACCCCAATGCGGCGAGCGAG 1020  
 Db 961 CAACCTCGCGCGAGATGAGAGGTCCTGCGCGCATCAACCCCAATGCGGCGAGCGAG 1020  
 QY 1021 CGCGGTGCTGACGCGGTGCGCTTGCACACCGGCGAGATGTGCTGTTGATGCGGCC 1080  
 Db 1021 CGCGGTGCTGACGCGGTGCGCTTGCACACCGGCGAGATGTGCTGTTGATGCGGCC 1080  
 QY 1081 TCGGCGCAATTCCTGTGGCGCGGTGATACCAACTACCAATATGATCGCTGATGAC 1140  
 Db 1081 TCGGCGCAATTCCTGTGGCGCGGTGATACCAACTACCAATATGATCGCTGATGAC 1140  
 QY 1141 GAGACCGGCTTGTGACGCGTGAACGAGATGCGGTGTTGAAAGGCTGACGTTGAATAT 1200  
 Db 1141 GAGACCGGCTTGTGACGCGTGAACGAGATGCGGTGTTGAAAGGCTGACGTTGAATAT 1200  
 QY 1201 GACGTCGCGCGACCTTCCTGGGTGGCGCGGACCTGTGTCAGCGCGCACTGAACCGGAC 1260  
 Db 1201 GACGTCGCGCGACCTTCCTGGGTGGCGCGGACCTGTGTCAGCGCGCACTGAACCGGAC 1260  
 QY 1261 ACCGCAATTTACTTCTTGGCGCTGAACAATGCTGCTACGATATATATGCGCGTATCAA 1320  
 Db 1261 ACCGCAATTTACTTCTTGGCGCTGAACAATGCTGCTACGATATATATGCGCGTATCAA 1320  
 QY 1321 GAGTTAGCGCGCTCGACGTTATTAACACGCGCGCAACCGCAAACTTCGCGCGGCGCTTT 1380  
 Db 1321 GAGTTAGCGCGCTCGACGTTATTAACACGCGCGCAACCGCAAACTTCGCGCGGCGCTTT 1380  
 QY 1381 GAAAAATATGCGCGCATGACGCGATTTGATATGACACCGGCGCGCACTTGTGTCGCGC 1440

```

Db      1381  GAAATATATGGGCGCATCGACGGATGATATCAGACCGGGCGCACCTTGTGTCGGCG 1440
QY      1441  GAGCGCCCTGGCGCGCAACTACTGCGCTTTGTGAGCGGCAAGCGGTGTGTTCAAC 1500
Db      1441  GAGCGCCCTGGCGCGCAACTACTGCGCTTTGTGAGCGGCAAGCGGTGTGTTCAAC 1500
QY      1501  GGGCGGACGACCGCTATTTCCTGTCCTCAGCCAGAGAAACCGCGAGACTTTGTGCGAG 1560
Db      1501  GGGCGGACGACCGCTATTTCCTGTCCTCAGCCAGAGAAACCGCGAGACTTTGTGCGAG 1560
QY      1561  GCGCGCTCTTCCAGCGGTGCGAGAGGGGCGAGGAGTACGATGAGTGGACCGCGTGA 1620
Db      1561  GCGCGCTCTTCCAGCGGTGCGAGAGGGGCGAGGAGTACGATGAGTGGACCGCGTGA 1620
QY      1621  TATATCGCATCGATCGGTCGGGCGGTCTGACCTATGAGCAGCAATTGAACCGCGCTG 1680
Db      1621  TATATCGCATCGATCGGTCGGGCGGTCTGACCTATGAGCAGCAATTGAACCGCGCTG 1680
QY      1681  GAGCGCATGATTCGACCTGCGGTGCGTATGCGATCTATGCTTTGCACTGCCGAGTAA 1740
Db      1681  GAGCGCATGATTCGACCTGCGGTGCGTATGCGATCTATGCTTTGCACTGCCGAGTAA 1740

RESULT 3
AD134117
ID      AD134117 standard; DNA; 1743 BP.
AC      AD134117;
XX
DT      15-APR-2004 (first entry)
XX
DE      Ketoglulonicigenium sp. sorbitol dehydrogenase (SDH) 3 DNA.
XX
KM      Sorbitol dehydrogenase; SDH; cytochrome c; molecular biology;
XX      bacteriology; industrial fermentation; L-sorbose; 2-keto-L-gulononic acid;
XX      2KLG; gene; ds.
XX
OS      Ketoglulonicigenium.
XX
FH      Key Location/Qualifiers
FT      CDS 1..1743
FT      /*tag= b
FT      /product= "SDH protein"
FT      sig_peptide 1..69
FT      /*tag= a
FT      mat_peptide 70..1740
FT      /*tag= c
FT      /product= "Mature SDH protein"
XX
XX      US2003228672-A1.
XX      11-DEC-2003.
XX
XX      06-JUN-2002; 2002US-00162713.
XX
XX      06-JUN-2002; 2002US-00162713.
XX
XX      (ARCH ) ARCHER-DANIELS MIDLAND CO.
XX
XX      Choi E, D'elia J, Kim H, Kim M, Lee UK, Pan J, Stoddard SF;
XX      Yum D;
XX      WPI; 2004-052025/05.
XX      P-PSDB; AD134121.
XX
XX      New proteins of sorbitol dehydrogenases and cytochrome c of the strains
XX      PT Ketoglulonicigenium spp.; useful in molecular biology, bacteriology and
XX      industrial fermentation, e.g. for producing L-sorbose and 2-keto-L-
XX      gulonic acid.
XX
XX      Claim 30; SEQ ID NO 4; 68pp; English.
XX

```

```

CC      The invention relates to the identification and isolation of nucleic acid
CC      sequences and proteins of sorbitol dehydrogenases (SDH) and cytochrome c
CC      of the strains, Ketoglulonicigenium sp. The proteins and nucleic acid
CC      molecules are useful in the fields of molecular biology, bacteriology and
CC      industrial fermentation specifically for producing L-sorbose and 2-keto-L-
CC      gulonic acid (2KLG). The present sequence is Ketoglulonicigenium sp. SDH
CC      DNA.
XX
SQ      Sequence 1743 BP; 361 A; 619 C; 496 G; 267 T; 0 U; 0 Other;
XX
Query Match 69.3%; Score 1206.2; DB 12; Length 1743;
Best Local Similarity 80.9%; Pred. No. 1e-292;
Matches 1406; Conservative 0; Mismatches 333; Indels 0; Gaps 0;
QY      1 ATGAACCCCAAAACGCTGCTTCCGACAGGCGCGCGTGTATGCTTACCGGCGCCG 60
Db      1 ATGCGACCCCAAAACGCTGCTTCCGACAGGCGCGCGTGTATGCTTACCGGCGCCG 60
QY      61 GCATTGCGGAGGTTAACCCCGATTAACCGATGAAGTCTGTGCGAACCAGCCGCGTGTAA 120
Db      61 GCCTTTGCGGAGTACCCCGATTAACCGATGAAGTCTGTGCGAACCAGCCGCGGCGAG 120
QY      121 TGGATTAACTAGCGCGCAACCAAGAAACTATGCGCACTGCGCCCTGACCCGATCACT 180
Db      121 TGGATTAACTAGCGCGCAACCAAGAAACTATGCGCACTGCGCCCTGAGAACGATTACG 180
QY      181 GCGCAACAAGTGTGCTCACTGTCAGACTGTGAGGCGCGGAGTGAAGGCGGCGCGTA 240
Db      181 ACCGACAACGTCGCGCAGCTGCACTGTGAGGCGCGGAGTGAAGGCGGCGCGTA 240
QY      241 CAGGTCAAGCGGATGATTCATGATGCGGTATGATTCATGCGCAACCCCGGTGATG 300
Db      241 CAGGTCAAGCGGATGATTCATGATGCGGTATGATTCATGCGCAACCCCGGTGATG 300
QY      301 CAGCGCTGATATGCGCAACAGCGCATCTGATCTGCGGACAGCGCGCGCACTGCGCG 360
Db      301 CAGCGCATGACCGCAAAACCGCGCATCTGATCTGCGGACAGCGCGCGCACTGCGCG 360
QY      361 GTCCGCAAGCTTAAACGCGCAAGCGCAACCGCGCGCGCTGCGCCCTTAAACGCAAG 420
Db      361 GTTGCCCTGCGTGAACGCGCAAGCGCAACCGCGCGCGCTGCGCCCTTAAACGCAAG 420
QY      421 CTCATTTCAGCTCATGCGCAACCATCTGATTTGCGGTGATTTGAGAGCGGCGCAGT 480
Db      421 CTCATTTCAGCTCATGCGCAACCATCTGATTTGCGGTGATTTGAGAGCGGCGCAGT 480
QY      481 GTATTGATGTCGAAACGTGATCGGCGGAGAGACGCGTTGACACAGTAAACACACAGG 540
Db      481 GTCTTTGATGTCGAGCGCGCGCTGCGGAGATGACGCGCTGACAGCAACACAGGCG 540
QY      541 ATTGTGCCCAATGCGCTCATGCTGTGCGGCTTCCACTGCGCAATTTGCTTATGATGC 600
Db      541 ATTGTGCCCAACGCGCTCATGCTGTGCGGCTTCCACTGCGCAATTTGCTTATGATGC 600
QY      601 TTATATCTCGGGGCGACATTCGCGGAGCGGTGAGAGCTGTGCGCAACCACTTTATCCG 660
Db      601 TTGTCTCGGGGTCAGACCGCGCGCGAGCGGAGAACTGTGCGGCAACTTCTTATCCG 660
QY      661 CAGCGGCGCAAGAGGAGTGAAGAGCTTGGGGCAATGATTTGAGGCGCGCTGATGAC 720
Db      661 CAAGCGGCGCAAGAGAGTGAAGAGCTTGGGGCAACTGATTTGAGGCGCGCTGATGAC 720
QY      721 GCGCTTGGGGTCAAGTCACTATGATCCGCTGACAGAACTTGTGTTCTATGCTGAC 780
Db      721 GCGCTTGGGGGCGAGTCACTATGATCCGCTGACAGAACTTGTGTTCTATGCTGAC 780
QY      781 GCGCTGAGGCCAGCGCTCGAACCAGACGCGGAGACGCGCGGCGCAAGCTGTATGCA 840
Db      781 GCGCTGAGGCCAGCGCTCGAACCAGACGCGGAGACGCGCGGCGCAAGCTGTATGCA 840
QY      841 AACACCGCTTTGCGGTGCTGCGGAGACGCGGAGAACTTGTGCGGCTGACAGACCTG 900
Db      841 AATACCGCTTTGCGGTGCTGCGGAGACGCGGAGAACTTGTGCGGCTGACAGACCTG 900

```

QY 901 CCGGCGGACAACTGGGACCAAGATGACGCTTCGAGATGATGTCGCCCAACGTCGATG 960  
 DB 901 CCGGCGGACAACTGGGACCAAGATGACGCTTCGAGATGATGTCGCCCAACGTCGATG 960  
 QY 961 CAACCCCTCGGCGGAGATGAGAGGCTGCGCGCATCAACCCCAATCGGCGAGCGGCGAG 1020  
 DB 961 CAGCGCGCTGCGGACATGAGCGGCTTCAGGCCATCAACCCCAACCGCAACGCGGAG 1020  
 QY 1021 CGCGGTGCTGACGGGTGCGCTTCGAGAGACCGGACGATGCTGCTTTGATGCGGCC 1080  
 DB 1021 CGTGGCTGCTGACGGGATCGCCCTCGCAAAACCGGACCGATGCTGCTTCGACGCGAA 1080  
 QY 1081 TCGGCGCAATTCCTGTTGGGCGCGGATACCAACTAACCAATATGATCGGCTCGATGAC 1140  
 DB 1081 ACCGCGCAATTCCTGTTGGGCGGACCGGACCACTAACCAACTGATCGGCTCGATGAC 1140  
 QY 1141 GAGACCGGCTTGTGACGCTGAAAGAGATGCGTGTGAAGAGCTGACGTTGAATAT 1200  
 DB 1141 GAAACCGGCTTGTGACGCTGAAAGAGATGCGTGTGAAGAGCTGACGTTGAATAT 1200  
 QY 1201 GAGCTGCGCGGACCTTCCTGGGTGGGCGGACGATGCTGACCGGACCTGAAACCGGAG 1260  
 DB 1201 GACATCTGCGCGACCTTCCTGGGAGCGGACGATGCTGCGGACCGCTGAAACCGGAG 1260  
 QY 1261 ACCGCGATTTACTTCTTGGCGCTGAACAATGCTGTACGATATATATGCGCGTGTATCA 1320  
 DB 1261 AGCGGATTTACTTCTTGGCGCTGAACAATGCTGTACGATATATATGCGCGTGTATCA 1320  
 QY 1321 GAGTTAGCGCGCTGACGCTTATACACAGCGGACCGCAAACTCGCGCGGCTTT 1380  
 DB 1321 GAAATTCGCGCGCTTGTATGTATCAATACAGCGGACCTTCAACAGCTTGCACCGGCTTT 1380  
 QY 1381 GAAATATGCGCGCATCGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
 DB 1381 GAAATATGCGCGCATCGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
 QY 1441 GAGCGCGCTGCGGAACTACTGCGCGCTTTGTGACGCGGACGCGTGTGTGATCAAC 1500  
 DB 1441 GAACTGCTGCGGAACTACTGCGCGCTTTGTGACGCGGACGCGTGTGTGATCAAC 1500  
 QY 1501 GCGCGGACCGGACGCTATTTCCGCGGCTTACGACGAGGAAACCGGCGGAGATTTGTGTGAG 1560  
 DB 1501 GCGCGGACCGGACGCTATTTCCGCGGCTTACGACGAGGAAACCGGCGGAGATTTGTGTGAG 1560  
 QY 1561 GCCCGTCTTGCACGCTGCGGACGCGGACGCGGATGATGATGATGATGATGATGATGATGAT 1620  
 DB 1561 ACCCGTCTTGCACGCTGCGGACGCGGACGCGGATGATGATGATGATGATGATGATGATGAT 1620  
 QY 1621 TATATGCCATCGGTGCGGCGGCTTCTGACTATGCGACGCAATTTGAACGCGCGCTGGCC 1680  
 DB 1621 TACGTGCGATCGGCGGCGGCGGACGACCTACGCGCAACCAACCAACCGTGGCTTACG 1680  
 QY 1681 GAGGCAATCATTCGACCTCGGTGCGGATGATGATGATGATGATGATGATGATGATGATGAT 1739  
 DB 1681 GAGGCAATCATTCGACCTCGGTGCGGATGATGATGATGATGATGATGATGATGATGATGAT 1739  
 RESULT 4  
 ADI34124  
 ID ADI34124 standard; DNA; 2281 BP.  
 XX  
 AC ADI34124;  
 DT 15-APR-2004 (first entry)  
 XX Ketogulonigenium sp. sorbitol dehydrogenase (SDH) 3 gene.  
 XX Sorbitol dehydrogenase; SDH; cytochrome c; molecular biology;  
 KW bacteriology; industrial fermentation; L-sorbose; 2-keto-L-gulononic acid;  
 KM 2KLG; gene; ds.  
 XX Ketogulonigenium.  
 OS

XX Key Location/Qualifiers  
 FH 521..524  
 FT RBS  
 FT  
 FT CDS  
 FT /tag= a  
 FT /standard\_name= "Shine-Dalgarno"  
 FT 530..2272  
 FT /tag= c  
 FT /product= "SDH protein"  
 FT 530..598  
 FT /tag= b  
 FT 599..2269  
 FT mat\_peptide  
 FT /tag= d  
 FT /product= "Mature SDH protein"  
 PN US2003228672-A1.  
 PD 11-DEC-2003.  
 PD  
 PD 06-JUN-2002; 2002US-00162713.  
 PF 06-JUN-2002; 2002US-00162713.  
 PF  
 PR 06-JUN-2002; 2002US-00162713.  
 PR  
 PA (ARCH ) ARCHER-DANIELS MIDLAND CO.  
 PI Choi E, D'elia J, Kim H, Kim M, Lee JK, Pan J, Stoddard SF,  
 PI Yum D;  
 PI Yum D;  
 DR MPI: 2004-052025/05.  
 DR P-PSDB; ADI34121.  
 DR  
 XX New proteins of sorbitol dehydrogenases and cytochrome c of the strains  
 PT Ketogulonigenium spp., useful in molecular biology, bacteriology and  
 PT industrial fermentation, e.g. for producing L-sorbose and 2-keto-L-  
 PT gulonic acid.  
 PS Claim 53; SEQ ID NO 11; 68bp; English.  
 PS  
 XX The invention relates to the identification and isolation of nucleic acid  
 CC sequences and proteins of sorbitol dehydrogenases (SDH) and cytochrome c  
 CC of the strains, Ketogulonigenium sp. The proteins and nucleic acid  
 CC molecules are useful in the fields of molecular biology, bacteriology and  
 CC industrial fermentation specifically for producing L-sorbose and 2-keto-L-  
 CC gulonic acid (2KLG). The present sequence is Ketogulonigenium sp. SDH  
 CC gene.  
 XX  
 XX Sequence 2281 BP; 483 A; 799 C; 632 G; 367 T; 0 U; 0 Other;  
 SQ  
 Query Match 69.3%; Score 1206.2; DB 12; Length 2281;  
 Best Local Similarity: 80.9%; Pred. No. 1.1e-292;  
 Matches 1406; Conservative 0; Mismatches 333; Indels 0; Gaps 0;  
 QY 1 ATGAACCCGACACGCTGCTTGTGACACGCGCGGCTGCTATTGCTTACCGCGCGGCC 60  
 DB 530 ATGCACCCGACACGCTGCTTGTGACACGCGCGGCTGCTATTGCTTGTGCGGATCCCT 589  
 QY 61 GGATTGCGGCGGATACCCCGGATTAACGATGAACTGCTGCGGAAACCGCGCGGTGTGAA 120  
 DB 590 GCTTTGCGGCGGATACCCCGGATTAACGATGAACTGCTGCGGAAACCGCGCGGTGTGAA 649  
 QY 121 TGGATTAACGCGCGGACCAACCAAACTATGCGCACTGCGCCCTGATACCGCAATCACT 180  
 DB 650 TGGATTAACGCGCGGACCAACCAAACTATGCGCACTGCGCCCTGATACCGCAATCACT 709  
 QY 181 GCCGACACGTTGCTCACTTGGCACTGCTGCGCGCGGCGGATGAGAGCGGCGCGCTA 240  
 DB 710 ACCGACACGTTGCTCACTTGGCACTGCTGCGCGCGGCGGATGAGAGCGGCGCGCTA 769  
 QY 241 CAGGTACGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
 DB 770 CAGGTACGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 829  
 QY 301 CAGGCGCTGATGCGCAACAGGCGATGATGATGATGATGATGATGATGATGATGATGATGAT 360

Db 830 CAGGCCATGACGCGCAAAACCGGCGACTGATGAGGAAACCGCCGCACTGCGGCC 889  
 Qy 361 GTGCCACGCTAAACCGCCGAAGCGACCGCAAGCGCGCTCGCCCTTTACGGCAGAGC 420  
 Db 890 GTTGCCTCGCTGACGCGCAAGCGACCGTTAAACGGCGCTCGCCCTTATGCGCAAC 949  
 Qy 421 CTGATTTAGCTCATGAGCAACCACTGATCGCGCTGATATGAGACGGCGCAGTC 480  
 Db 950 CTCTATTTACCTCGTGGCAACACACTTGTGCACTGGACATGGCCACCGCCCAAGTC 1009  
 Qy 481 GTATTCGATGTCGACGTGATCGGCGAAGACCGCTTGAACAGTAACACACGCGGCCG 540  
 Db 1010 GTCTTGTATGTCGACGCGCGCTCGGCGATGACGCGCTGACGCAACACACGCGGCCG 1069  
 Qy 541 ATTGTGCGCAATGCGCTCATCGTGGCGGTTCCACCTGCGCAATATTCGCGCTATGATG 600  
 Db 1070 ATTGTGCGCAACGCGCTCATCGTGGCGGTTCCACCTGCGCAATATTCGCGCTATGATG 1129  
 Qy 601 TTTATTCGCGGAGCAGATTCGCGGACGAGTGAAGCTGTGGCGCAACCACTTTATCCCG 660  
 Db 1130 TTTGTCTCGGGTACGACCGCGCCAGCGGGAAGAACTGTGGCGCAACTTCACTTCCCG 1189  
 Qy 661 CAGCGCGCGAAGAGGTGACGAGACTTGGGCGCAATGTTTGAAGCGCGCTGAGATGACC 720  
 Db 1190 CAGCGCGCGAAGAGGTGACGAGACTTGGGCGCAATGTTTGAATCGCGCTGAGATGACC 1249  
 Qy 721 GGGGTCTGGGTCAGATCACTTATGATCCCGTACGCAACCTTGTATGAGCTCGACC 780  
 Db 1250 GGGGTCTGGGTCAGATCACTTATGATCCCGTACGCAACCTTGTATGAGCTCGACC 1309  
 Qy 781 GGGGTGGGCCCCAGCTTCCAAACCCGCGGACCGCGCGCGCGCTGATATGCGACC 840  
 Db 1310 GGGGTGGGCCCCAGCTTCCAAACCCGCGGACCGCGCGCGCGCTGATATGCGACC 1369  
 Qy 841 AACACCGCTTTCGCGTCCCGCACGCGGCGAATGTCGCGGTGACGACGACCTG 900  
 Db 1370 AATACCGCTTTCGCGTCCCGCACGCGGCGAATGTCGCGGTGACGACGACCTG 1429  
 Qy 901 CCGCGCGCACTGCGGACCAAGATGCACTGATGATGTCGCGCACTCGATGTCG 960  
 Db 1430 CCGCGCGCACTGCGGACCAAGATGCACTGATGATGTCGCGCACTCGATGTCG 1489  
 Qy 961 CAACCCCTCGCGGATGAGGCGTGGCGCGCATCAACCCCAATCGCGCGACGCGGCGAG 1020  
 Db 1490 CAGCGCGCTGCGCATGACGCGCTTCAAGCGCATCAACCCCAATCGCGCGACGCGGCGAG 1549  
 Qy 1021 CCGCGTGTCTGACGCGGTGCGCTTGCAGACCGGCAAGATGTCGCTTATGTCGCGCC 1080  
 Db 1550 CGTGCCTGTCTGACGCGGATCCCTGCAAAACCGGCAAGATGTCGCTTATGTCGCGCC 1609  
 Qy 1081 TCGCGGCAATTCCTGTCGCGGCGTGAATCAACTACCAATATGATGCGCTCGATCGAC 1140  
 Db 1610 ACCGCGCAATTCCTGTCGCGGCGTGAATCAACTACCAATATGATGCGCTCGATCGAC 1669  
 Qy 1141 GAGACCGGCTTGTGACGCTGAACGAGATGCGGTCGTAAGAGCTGACGCTTGAATAT 1200  
 Db 1670 GAAACCGGCTTGTGACGCTGAACGAGATGCGGTCGTAAGAGCTGACGCTTGAATAT 1729  
 Qy 1201 GAGCTCTGCGCGACCTTCTCGGCTGCGGCGTGAATGTCGTCGACCGCGCACTGAACCGGAC 1260  
 Db 1730 GACATCTGCGCGACCTTCTCGGCTGCGGCGTGAATGTCGTCGACCGCGCACTGAACCGGAC 1789  
 Qy 1261 ACCGCGATTTACTTCTTGCCTGGAACCAATGCTGCTAGCAATATATGCGCGCTTGAATCA 1320  
 Db 1790 AGGCGATCTACTTACTTACCGCTGAAACACCGCTGCTGAGATCACTAGCGCGCTTGAATCA 1849  
 Qy 1321 GAGTTTACGCGCTCGACGCTTATTAACACGAGCGCACTGCAAAATTCGCGCGCGCTT 1380  
 Db 1850 GAAATTCGCGCGCTTATGATGTAACATCAAGCGCACTTCAACAGCTTACACCGCGCTT 1909  
 Qy 1381 GAAATATGCGCGCGCATGACGCGATGATATCAGACACGCGCGCGCACTTGTGTCGCGG 1440  
 Db 1910 GAAAAATGCGCGCGCATGACGCGATGATATCAGACACGCGCGCAAACTCTGTGTGTCGCGG 1969

Qy 1441 GAGCGCCCTGCGCGGAACTACTGCGCGCTTTTGTGACGCGGCGGCTGTGATTCAC 1500  
 Db 1970 GAACTCTGCGCTGAACTACTGCGCGCTTCTCTGACGCGCTGCGCGCTGTCTTCAAC 2029  
 Qy 1501 GCGCGAACCAGCCGCTATTTCCGTGCGCTCAGACCGGAAACCGGCGAGACTTTGTGGCAG 1560  
 Db 2030 GCGCGAACCAGCCGCTATTTCCGTGCGCTCAGACCGGAAACCGGCGAGACTTTGTGGCAG 2089  
 Qy 1561 GCGCGTCTTGGAGAGGTGCGGACCGGCGGCGGCGATCAAGAGTGTGACGCGGTCGAA 1620  
 Db 2090 ACCGCTCTGCGAGATGTGCTTACCGGCGAAGCACTACGAAATCGACGGGACCCAA 2149  
 Qy 1621 TATATCGCATCGGTGCGGCGGTCTGACCTATGCGCAAGCAATGAAACGCGCGCTGAGCC 1680  
 Db 2150 TACGTCGAGATCGCGGGGGGGGCGACACTTACGGGCAACCAACCAACCGTCCCTGAGC 2209  
 Qy 1681 GAGGCATTCGATTTGACCTCGGTGCGTAATGCGATCTATGCTTTGTGACATGCGCGGCTA 1739  
 Db 2210 GAGCGATCGACTGACCAAGATCGGCAACGCGGTTTACGTTTGTGCGTGGCGACGA 2268

RESULT 5  
 AAV29051  
 ID AAV29051 standard; CDNA; 1740 BP.  
 XX  
 AC AAV29051;  
 XX  
 DT 21-AUG-1998 (first entry)  
 XX  
 DE Alcohol and/or aldehyde dehydrogenase A nucleotide sequence.  
 XX  
 KM Alcohol/aldehyde dehydrogenase A enzyme; recombinant organism; aldehyde;  
 KM ketone; carboxylic acid; L-sorbose; D-sorbitol; 2-keto-L-gulononic acid;  
 KM L-ascorbic; inhibition; de.  
 XX  
 OS Gluconobacter oxydans.  
 XX  
 FH Key Location/Qualifiers  
 FT 1..1737  
 FT CDS /product= "alcohol and/or aldehyde dehydrogenase A"  
 PN EP832974-A2.  
 PD 01-APR-1998.  
 XX  
 PF 11-SEP-1997; 97EP-00115801.  
 XX  
 PR 19-SEP-1996; 96EP-00115001.  
 XX  
 PA (HOFF) HOFFMANN LA ROCHE & CO AG F.  
 PI Asakura A, Hoshino T, Ojima S, Shinjoh M, Tomiyama N;  
 DR WPI; 1998-195228/18.  
 DR P-PSDB; AAM37873.  
 XX  
 PT Recombinant Gluconobacter oxydans alcohol and/or aldehyde dehydrogenase  
 PT enzyme(s) - useful for converting substrate(s), e.g. L-sorbose or D-  
 PT sorbitol to 2-keto-L-gulononic acid.  
 XX  
 PS Disclosure; Page 27-28; 59pp; English.  
 XX  
 CC This is the nucleotide sequence encoding the Gluconobacter oxydans  
 CC alcohol and/or aldehyde dehydrogenase A enzyme. The enzymes or  
 CC recombinant organisms can be used to convert suitable substrates to  
 CC aldehydes, ketones or carboxylic acids, especially to convert L-sorbose  
 CC or D-sorbitol to 2-keto-L-gulononic acid, which can be converted to L-  
 CC ascorbic acid by standard procedures. The derivatives of ADH enzymes  
 CC have desired substrate specificity, higher affinity to a substrate, lower  
 CC affinity to an inhibitory compound, higher stability against temperature  
 CC and/or pH and higher catalytic speed

XX Sequence 1740 BP; 356 A; 558 C; 507 G; 319 T; 0 U; 0 Other;  
 SQ Query Match 65.0%; Score 1130.2; DB 2; Length 1740;  
 Best Local Similarity 78.8%; Pred. No. 1.3e-273;  
 Matches 1374; Conservative 0; Mismatches 363; Indels 6; Gaps 2;

QY 1 ATGAACCCCAACAGCTGCTTCCAGCAGCGCGCGCTGCTATTGCTTACCGCGCCGCC 60  
 DB 1 ATGAACCCAGCTTGGCTTGGGCGAGTGTGGGCACTTGCAATGCTGCGGCAACC 60

QY 61 GCATTGCGGAGGTAACCCCGATTACCGATGAACCTGTGCGCAACCCCGCCGTGTGA 120  
 DB 61 GCCTTGTCTAAATGACCCCGTACCGATGATTTGTGCGCAACCCCGCGCTGTGA 120

QY 121 TGGATTACTACGCGCGCAACAAGAAATATATGCGCACTCGCCCTGACCCAGATCAT 180  
 DB 121 TGGATCAGCTACGCTGAGAACCAAGAAATTAACCTGCACTCGCCCTGACCGATCAG 180

QY 181 GCCGACAAGCTGTGATGCTGCACTGTGCACTGTGCGCGCGGATGAGAGCGGGGCGTA 240  
 DB 181 ACTGAGAACCTCGCGCAACTGCAACTGTGCTGCGCGCGCGCATGCAAGCGCGCAAGTC 240

QY 241 CAGGTACCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300  
 DB 241 CAGGTACCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300

QY 301 CAGGCGCTGATGCGCAACAGGCGATCTGATCTGGAACACCGCGCGCAACTGCGCGC 360  
 DB 301 CAGGCGCTGACGCAAAACTGGCGATCTGATCTGGGAACACCGCGCGCAACTGCGGAA 360

QY 361 GTGCGCCAGCTTAAAGCGCGCAAGCGGACCGGCGCGCTCGCCCTTTTACGCGAGAC 420  
 DB 361 ATGCGCAAGCTGAAACAGCTTTGGGAGCGGCGCGCATGGCGGTGACGCGCAACAC 420

QY 421 CTCTATTCTCATGAGGACAACCATCTGATCGCGCTGATATGAGACGCGCGCAGTTC 480  
 DB 421 GTTACTCTTGTTCGTGGGCAACACACTGTGCGCTTGAACCGCAACTGCGCGCAAGT 480

QY 481 GTATTGATTCGAACTGTGATCGGGGAAAGACGCGCTTGAACAGTAAACACACGCGG 540  
 DB 481 ACCTGACGCTGACCGCGCGCAAGGGAAGACATGATGCTTCG---AACTGTCGCGCGCG 537

QY 541 ATTGCGCGCATGCGCGTCACTGTCGCGGCTTCAACCTGCAATATTCGCGCTATGATG 600  
 DB 538 ATGCGCGCAACGCGCTGATGCTGTCGCGGCTTCACTGCGCAATATTCGCGCTATG 597

QY 601 TTTATCTCGGGGACGATTCGCGGACGCGGTGAGAGCTGTGCGCAACCATTTATCCG 660  
 DB 598 TTTGTCGCGGCGACGACTCGGCGCACCGGTGAAGAGCTGTGCGCAACTTCACTCCG 657

QY 661 CAGCGCGCGCAAGAGGTGACGAGCTTGGGGCAATGATTTGAGGCGCGCTGAGTAC 720  
 DB 658 CGCGCTGCGCAAGAGGTGATGAGCTTGGGCAAGATTAAGAACCCGCTGTGAGTAC 717

QY 721 GCGGCTGCGGCTGATCACTATGATCCCGTGAACCACTTGTGTTCTATGCTGAC 780  
 DB 718 GGTGCTGCGGCGCAAGTCACTATGACCCCGTCAACCACTTGTCACTACGCTGAC 777

QY 781 GCGGCGCGCGCAAGCTTCCGAAACCCAGCGCGGACCGCGCGCGCATGATGAGAC 840  
 DB 778 GCTGTGGGTCCGCGTCCGAAACCCAGCGCGCACCCCGCGCGCGCATGATGAGAC 837

QY 841 AACACCGGCTTTGCGGCTGCTCCGACACGCGCGAGATTTGTCGCGTCAACGACCTG 900  
 DB 838 AACACCGGCTTTGCGGCTGCTCCGACACGCGCGAGATTTGTCGCGTCAACGACCTG 897

QY 901 CCGCGCGCAACTGCGGACCAAGATGCACTTGAAGATGATGTCGCGCAAGTCAAGT 960  
 DB 898 CCGCGCGCAACTGCGGACCAAGATGCACTTGAAGATGATGTCGCGCAAGTCAAGT 957

QY 961 CAACCCCTCGGCGAGATGAGAGGTGCGCGCATCAACCCCAATGCGCGGACGCGCGAG 1020

DB 958 CAACCCCTGACGAGATGAGAAAGTCTGCACTGATCAACCCGAAACGCCCACTGCGAG 1017  
 QY 1021 CGCGGTGCTGAGCGGCTGCGCTTGAAGAACCGGCAACGATGATGATGATGATGATGATG 1080  
 DB 1018 GGTGCGGCTGACCGGCGGCTTCCGTAACCGGCAACGATGATGATGATGATGATGATG 1077

QY 1081 TCGGCGGAATTCCTGTGGGCGCGGATGATCAACTACCAATATGATGATGATGATGATG 1140  
 DB 1078 ACCGCGAATTCCTGTGGGCGCGGATGATCAACTACCAATATGATGATGATGATGATG 1137

QY 1141 GAGACCGGCTTGTGACGCTGATACGAGATGCGGTGCTGAAAGCTGGAAGCTTGAAT 1200  
 DB 1138 GAAACCGGCACTGTACCGGTAACGAAAGATGATGATGATGATGATGATGATGATG 1197

QY 1201 GACGTCTCCGACCTTCTGCGGTGCGGCGCACTGATGATGATGATGATGATGATGATG 1260  
 DB 1198 GACGTCTCCGACCTTCTGCGGTGCGGCGCACTGATGATGATGATGATGATGATGATG 1257

QY 1261 ACCGCGAATTCCTGTGGGCGCGGATGATCAACTACCAATATGATGATGATGATGATG 1320  
 DB 1258 AGCGGCACTGATCTTCACTGATGATGATGATGATGATGATGATGATGATGATGATG 1317

QY 1321 GAGTTAGCGGCTGACGCTGATTAACACGAGCGGACCGGAAACCTGCGCGGCTTT 1380  
 DB 1318 GATTTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1377

QY 1381 GAAATATGAGCGCGCATGACGCGATGATGATGATGATGATGATGATGATGATGATG 1440  
 DB 1378 GATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1437

QY 1441 GAGCGCGCTGCGGCGAATCTGCGCGCTTGTGCGAGCGGAGCGGATGATGATGATG 1500  
 DB 1438 GAACTGCTGCGGCGAATCTGCGCGCTTGTGCGAGCGGAGCGGATGATGATGATG 1497

QY 1501 GCGCGGACGACCGGCTATTTCCGTCGCTCAGCGGAAACCGGCGAGCTTGTGCGAG 1560  
 DB 1498 GGTGTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1557

QY 1561 GCGCGCTTGTGCGAGGCTGCGGACGCGGAGCGGATGATGATGATGATGATGATGATG 1620  
 DB 1558 ACCGCGCTTGTGAAACGCTGCGGCGGCGGCGGATGATGATGATGATGATGATG 1617

QY 1621 TATATCGCATGCGGCGGCGGCTGATGATGATGATGATGATGATGATGATGATGATG 1679  
 DB 1618 TATGTCGATGCGGAGGCTGATGATGATGATGATGATGATGATGATGATGATG 1677

QY 1680 --CGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1737  
 DB 1678 GCGGAGCGAGTCACTGACCGCGCATGCGTAAACCGCGCTGATGATGATGATG 1737

QY 1738 TPA 1740  
 DB 1738 TPA 1740

RESULT 6  
 AAV29053  
 ID AAV29053 standard; cDNA; 1737 BP.

XX AAV29053;  
 DT 10-AUG-1998 (first entry)  
 XX  
 DE Alcohol and/or aldehyde dehydrogenase A', nucleotide sequence.  
 XX  
 KW Alcohol/aldehyde dehydrogenase A', enzyme; recombinant organism;  
 KW aldehyde; ketone; carboxylic acid; L-sorbose; D-sorbitol;  
 KW 2-keto-L-gulononic acid; L-ascorbic; inhibition; ds.  
 OS Gluconobacter oxydans.  
 XX  
 Key Location/Qualifiers  
 FH 1.1737  
 FT CDS

/\*tag= a  
/product= "alcohol and/or aldehyde dehydrogenase A'"

EP832974-A2.  
01-APR-1998.

11-SEP-1997; 97EP-00115801.

19-SEP-1996; 96EP-00115001.

(HOFF) HOFFMANN LA ROCHE & CO AG F.

Aakura A, Hoshino T, Ojima S, Shinjoh M, Tomiyama N;

WPI, 1998-195228/18.

P-PSDB; AAM37875.

Recombinant Gluconobacter oxydans alcohol and/or aldehyde dehydrogenase  
PT enzyme(s) - useful for converting substrate(s), e.g. L-sorbose or D-  
sorbitol to 2-keto-L-gulonic acid.

Disclosure; Page 31-32; 59pp; English.

This is the nucleotide sequence encoding the Gluconobacter oxydans  
alcohol and/or aldehyde dehydrogenase A' enzyme. The enzymes or  
recombinant organisms can be used to convert suitable substrates to  
aldehydes, ketones or carboxylic acids, especially to convert L-sorbose  
or D-sorbitol to 2-keto-L-gulonic acid, which can be converted to L-  
ascorbic acid by standard procedures. The derivatives of ADH enzymes  
have desired substrate specificity, higher affinity to a substrate, lower  
affinity to an inhibitory compound, higher stability against temperature  
and/or pH and higher catalytic speed

Sequence 1737 BP; 339 A; 570 C; 509 G; 319 T; 0 U; 0 Other;

Query Match 62.1%; Score 1079.8; DB 2; Length 1737;

Best Local Similarity 77.1%; Pred. No. 5,86-261;

Matches 1341; Conservative 0; Mismatches 392; Indels 6; Gaps 2;

QY 1 ATGACCCCAACGCTGCTTGCACAGGCGCGCTGTATGCTTACCGCCGCC 60  
DB 1 ATGAACTGACGACCTGCTGCAAGCAGCGCCCTGCTGTGTTGGACACATCCC 60  
QY 61 GCATTGCGCGAGTAACCCCGATTACCGATGAATCTGCTGCGAACCCTGCTGTGA 120  
DB 61 GCCCTTGCCCA---AACCGCATCACCGATGAATCTGCTGCGAACCCTGCTGTGA 117  
QY 121 TGGATTAACTAGCGCGCAACCAAGAACTATCGCACTCGCCCTGACCCAGATCACT 180  
DB 118 TGGATCACTAGCGCTGCAAGCAAGAACTATCGCACTCGCCCTGACCGATTAAC 177  
QY 181 GCCGACACGCTTGCTCACTGCTGCGCCCGCGGATGAGAGCGGCGCGTA 240  
DB 178 GCAGACACGCTGCGCAACTGCACTGCTGCGCGCGGTATGGAACGCGCAAGATC 237  
QY 241 CAGGTACCGCGCATGATTCATGATGCGCTGATATCTGCGCAACCCCGCTGATGATC 300  
DB 238 CAGGTACCGCGCTTGTGCTGATGCGCGTATATCTGCGCAACCCCGCTGATGATC 297  
QY 301 CAGGCGCTGATGCGCAACAGCGCATCTGATCTGCGAAACACCGCGCCACTGCGCGC 360  
DB 298 CAGGCGCATGACCGCGCGACCGCGCATCTGATCTGCGAAACACCGCGCAAC 357  
QY 361 GTGCGCACGCTAAACCGCGCGACCGCGCGCGCTGCGCTTTACGCGACGAC 420  
DB 358 ATGCGCACGCTGAACGCTTGTGATGCGCGACCGCGCGCATGCGCTTATGCGCAAC 417  
QY 421 CTGATTTTACGCTCATGCGCAACCACTGATGCGCTGATATGAGACGCGCGACGCTC 480  
DB 418 GTCTATTTTCTCTGCGCAACCACTGATGCGCTGATGACACCTGCGCAACGCGCAAGTC 477  
QY 481 GTATTGATGTGACGCTGATGCGCGCAACGCGCTTGAACGATTAACCAACGCGCGCG 540

DB 478 GTATTGACGCTGATTCGCGGTCAAGGACCGAAT---TGGTCTCGAAGCTGTGCGCGCG 534  
QY 541 ATTGTGCGCAATGCGCTGATGCTGCGGGTTTCACTGCGCAATTTGCGCTTATGATGC 600  
DB 535 ATTGTGCGCAATGCGCTGATGCTGCGGGTTTCACTGCGCAATTTGCGCTTATGATGC 594  
QY 601 TTTATCTCGGGGCGACGATTCGCGGAGGAGAGTGTGCGCGCAACCTTTATCCG 660  
DB 595 TTGCTTTGCGCGCACGATCTGCGCGACCGGTGAAGAGCTGTGCGCGCAACCTTTATCCG 654  
QY 661 CAGCGCGCGCAAGAGGAGTGAAGACTTGGGGCAATGATTTGAGCGCGCTGATGAC 720  
DB 655 CGCGCGCGCAAGAGGAGTGAAGACTTGGGGCAATGATTTGAGCGCGCTGATGAC 714  
QY 721 GCGCTGCGGGTCAAGTCACTGATGATCCCGTGAAGAACTTGTGTTATGAGCTGAC 780  
DB 715 GCGCTTTGCGCGCAAGTCACTGATGATCCCGTGAAGAACTTGTGTTATGAGCTGAC 774  
QY 781 GCGCTGCGGGTCAAGTCACTGATGATCCCGTGAAGAACTTGTGTTATGAGCTGAC 840  
DB 775 GCAATTGCGCGCTGCGCGCGATTCAGCGCGCACCGTTGCGCGCTGATGATGAC 834  
QY 841 AACCGCGCTTTGCGCGCTGCGCGCGATTCAGCGCGCACCGTTGCGCGCTGATGAC 900  
DB 835 AACCGCGCTTTGCGCGCTGCGCGCGATTCAGCGCGCACCGTTGCGCGCTGATGAC 894  
QY 901 CCGCGCGCAACTGCGCGCAAGATGACGCTTGAAGTATGCTGCGCAACGCTGATG 960  
DB 895 CCGCGCGCAACTGCGCGCAAGATGACGCTTGAAGTATGCTGCGCAACGCTGATG 954  
QY 961 CAACCTTGGCGCGAGATGAGAGGCTGCGCGCATCAACCCCAATGCGCGAGCGGCGAG 1020  
DB 955 CAGCGCTGCGCTGAGTGAAGAGGCTGCGCGCATCAACCCCAATGCGCGAGCGGCGAG 1014  
QY 1021 CGCGCTGCGCTGAGTGAAGAGGCTGCGCGCATCAACCCCAATGCGCGAGCGGCGAG 1080  
DB 1015 CTTGCGCTTGAAGAGGCTGCGCGCATCAACCCCAATGCGCGAGCGGCGAG 1074  
QY 1081 TGGCGCAATTCCTGCGCGCGCGCTGATTAACCACTAATGATGCTGCTGATGAC 1140  
DB 1075 ACCGCGCAATTCCTGCGCGCGCGCTGATTAACCACTAATGATGCTGCTGATGAC 1134  
QY 1141 GAGACCGCGCTGAGTGAAGAGGCTGCGCGCATCAACCACTAATGATGCTGCTGATGAC 1200  
DB 1135 CCGATGCTGCGTGAAGAGGCTGCGCGCATCAACCACTAATGATGCTGCTGATGAC 1194  
QY 1201 GACGCTGCGCGCTGCGCGCTGCGCGCATCAACCACTAATGATGCTGCTGATGAC 1260  
DB 1195 GAATTCGCGCGCTGCGCGCTGCGCGCATCAACCACTAATGATGCTGCTGATGAC 1254  
QY 1261 ACCGCGCATTTACTTCTGCGCGCTGCGCGCATCAACCACTAATGATGCTGCTGATGAC 1320  
DB 1255 ACTGCGCATTTACTTCTGCGCGCTGCGCGCATCAACCACTAATGATGCTGCTGATGAC 1314  
QY 1321 GAGTTTACGCGCTGCGCGCTGCGCGCATCAACCACTAATGATGCTGCTGATGAC 1380  
DB 1315 GAGTTTACGCGCTGCGCGCTGCGCGCATCAACCACTAATGATGCTGCTGATGAC 1374  
QY 1381 GAAATATGCGCGCGCATCAACCACTAATGATGCTGCGCGCATCAACCACTAATGATGAC 1440  
DB 1375 GAAATATGCGCGCGCATCAACCACTAATGATGCTGCGCGCATCAACCACTAATGATGAC 1434  
QY 1441 GAGCGCGCTGCGCGCATCAACCACTAATGATGCTGCGCGCATCAACCACTAATGATGAC 1500  
DB 1435 GAGCGCGCTGCGCGCATCAACCACTAATGATGCTGCGCGCATCAACCACTAATGATGAC 1494  
QY 1501 GCGCGGACCGACCGCTAATTTCTGCGCGCATCAACCACTAATGATGCTGCGCGCAT 1560  
DB 1495 GCGCGGACCGACCGCTAATTTCTGCGCGCATCAACCACTAATGATGCTGCGCGCAT 1554  
QY 1561 GCCCGTCTTGGAGCGCTGCGCGCATCAACCACTAATGATGCTGCGCGCATCAACCACT 1620





Qy	1081	TCGGGCGAATTCCTGTGGGCGCGGTGATGACCACTACACACATATATGATCGCCTCATGCAC	1140
Db	1078	ACGGGCGCAATTCGTGTGGGCGCGGACACCAACTACCAAAACATGATCAGTTGCATGCAC	1137
Qy	1141	GAGACCGGCGCTTGTGACGGGTACAGAGATGCGGTGCTGAAGAGCGTGAAGCTTGATAT	1200
Db	1138	GAACCGGTCGTGTACGGGTAAATGAAGATATCACTCTAAAAGATCTGGACACCGACATAC	1197
Qy	1201	GACGTCGCGCCGACCTTCTGGGTGGGCGCGCACTGATGTCAGCGCGCACTGAACCCGAGAC	1260
Db	1198	CGCATTTGCCCGGACATTTCTTGGTGGAGCGCACTGCGCGGTGATCCTTGAACCCGAT	1257
Qy	1261	ACGGGCAATTAATCTTCTGGCGGTAAACAATGCGCTGATACGATATATATGCGCGTGAATCA	1320
Db	1258	AGCGGCATCTAATCTCATTTCCCTGAACACAGCGCTGTGGGATTTGGCGGACGTGATCA	1317
Qy	1321	GAGTTTACCGCGCTGACGTCGTATTAACAACAGCGCGCACCGCAAAACTGCGCGCGGCTTT	1380
Db	1318	GAGTTTACCGGCATATGACGTCCTAACACACCGCGCACTTACCTGCTTGGCCCGGAAAA	1377
Qy	1381	GAATATATGGCGCGCATGACGCGATTTGATATGACGACCGGCGCGCACTTGTGTGCGCG	1440
Db	1378	GAATATATGGCGCGCATGACGCGATGACGACATGACGACGCGCAAAACCTGTGTGCGCTC	1437
Qy	1441	GAGCGCCCTGCGCGCAACTACTGCGCGCTTTGTGACGCGCGCGGTGTGTGTTCAAC	1500
Db	1438	GAAGTCTGCGGTGGAATCTATCTGCGCGGTCTCTGACGCGGTGCGCGCGTGTGTTCAAC	1497
Qy	1501	GAGCGGACCGACCGCTATTTCCGTGCGCCTCAGCGAGAAACCGGCGGAGACTTTGTGCAG	1560
Db	1498	GCGCGGACGCAATCGCTACTTCCGTGCGCCTCAGCGAGAAACTGGCGAGAACCTGTGGCAG	1557
Qy	1561	GCGCGTCTTGGGACGCTGTGCGACGCGGCGAGCGCATCACTTACGATTTGAGAGGGGTGCA	1620
Db	1558	ACCGGTCGCGGACGTGTGCGCACGCGGTAAACCATCACTACGACTGGAAGGCGGTGCA	1617
Qy	1621	TATATCGCATCGGTGCGGCGCGTCTGACTCATATG3ACAAGCAATTGAACGCGCGCCTGAC	1680
Db	1618	TATGTTCGCAATCGACGCGGCGGTAAATCTTAACG3CACTAACGTAACAGCAATATG3GC	1677
Qy	1681	GAGGCATTCGATTCGACTCGGTGTGGTAAATGCAATTCATGTCTTTGACATGCGCGGACATA	1740
Db	1678	GCGACCATTCGATTCGATTCGATTCG3GACCGCGGTCTTACGCTTCGCGCCTTCCGCAATTA	1737

RESULT	8
AD134122	
ID	AD134122 standard; DNA; 2519 BP.
XX	
AC	AD134122;
XX	
DT	15-Apr-2004 (first entry)
XX	
DE	Ketogulonidigenium sp. sorbitol dehydrogenase (SDH) 1 gene.
XX	
KM	Sorbitol dehydrogenase; SDH; cytochrome c; molecular biology
KM	bacteriology; industrial fermentation; L-sorbose; 2-keto-L-gu
XX	2KUG; gene; ds.
XX	
OS	Ketogulonidigenium.
XX	
FH	Location/Qualifiers
FT	RBS
FT	741..744
FT	/**tag= a
FT	/standard_name= "Shine-Dalgarno"
FT	750..2486
FT	/**tag= c
FT	/product= "SDH protein"
FT	750..818
FT	/**tag= b
FT	819..2483
FT	/**tag= d
FT	/product= "Mature SDH protein"

XX US2003228672-A1.  
 XX  
 XX 11-DEC-2003.  
 XX  
 XX 06-JUN-2002; 2002US-00162713.  
 XX  
 XX 06-JUN-2002; 2002US-00162713.  
 XX  
 XX (ARCH) ARCHER-DANIELS MIDLAND CO.  
 XX  
 XX Choi E, D'elia J, Kim H, Kim M, Lee JK, Pan J, Stoddard SF,  
 XX Yum D;  
 XX WPI; 2004-052025/05.  
 XX P-PSDB; ADI34118.  
 XX  
 XX New proteins of sorbitol dehydrogenases and cytochrome c of the strains  
 PT *Ketogulonigenium* spp., useful in molecular biology, bacteriology and  
 PT industrial fermentation, e.g. for producing L-sorbose and 2-keto-L-  
 PT gulonic acid.  
 XX  
 XX Claim 37; SEQ ID NO 9; 68bp; English.  
 XX  
 XX The invention relates to the identification and isolation of nucleic acid  
 CC sequences and proteins of sorbitol dehydrogenases (SDH) and cytochrome c  
 CC of the strains, *Ketogulonigenium* sp. The proteins and nucleic acid  
 CC molecules are useful in the fields of molecular biology, bacteriology and  
 CC industrial fermentation specifically for producing L-sorbose and 2-keto-L-  
 CC gulonic acid (2KLG). The present sequence is *Ketogulonigenium* sp. SDH  
 CC gene.  
 CC  
 CC Sequence 2519 BP, 542 A; 797 C; 703 G; 477 T; 0 U; 0 Other;

Query Match	Best Local Similarity	51.6%	Score 1071.2	DB 12	Length 2519
Query Match	Similarity	76.4%	Pred. No. 9.3e-259		
Matches 1329	Conservative	0	Mismatches 408	Indels	3
				Gaps	1
Qy	1	ATGAAACCCCAACAAAGCTGCTTTCGACACAGCGCGCGCTGCTATTGCTTACCGCGCCGCC	60		
Db	750	ATGAATTCGAATTGTTGCTTTCGCAAGAGTTCCTCCGTCGATCTTTCGTCGTGCC	809		
Qy	61	GCATTTCGCGAGGTAAACCCCGATTACCGATGAACTGTGCGGAACCCGCGCGTGTGAA	120		
Db	810	GCATTTCGCGATGTGAAGCCCGCTACCGACGAGTGTATGAAACCCGCGCGCGGAGA	869		
Qy	121	TGATTTACTACGAGCGCGCAACCAAGAAACTATGCGCACTCGCCCTGACCACATCACT	180		
Db	870	TGATTCAGCTATGAGCCCGCAACCAAGAAACTACCGCACTCGCCGCTGAACCAATACCC	929		
Qy	181	GCCGACAAAGTTGGTCACTTTCGCACTGTGCTGCGCGCGCGAGATGAGAGCGGGGCGCTA	240		
Db	930	CCCGACAAAGTCGCGCCAGCTGACCTGATCTGGCGCGCGCGGATGAACCCCGCGTGTG	989		
Qy	241	CAGGTACCGCCGATGATTCATGATGAGGGGTATGATCTGGCAACCCCGGATGATCATC	300		
Db	990	CAGGTGACCCCGCTGATTCACGAGGGGTATGATCTGGGGAACCCAGGCGATCATATT	1049		
Qy	301	CAGGCGCTGATGTCGGCAAAACAGCGCATCTGATCTGGCAAAACCGCGCCAACTGCCGCC	360		
Db	1050	CAGCGATTGACCGCAAAAACCGGTGACTTATCTGGGAACACCGCGCCAACTGCCGAG	1109		
Qy	361	GTCGCCACGCTAAACGCCCAAGCGGACCGCAAGCGCGCGCTCGCCCTTTACGTCACGAC	420		
Db	1110	ACCTCGACGCTCAGCTCGCTGGGGGATTCGCAAGCGCGGCACTCGGCTTTTATGCAACAA	1169		
Qy	421	CTCATTTGACGTCAATGGGCAACCATCTGATCGCGTGTGATATGAGAGCGGGCACGTC	480		
Db	1170	GTCATCTTCGTCCTGTGGGCAACCATATGATCGCGCTGATGCTGCCAGGGGCAATCTC	1229		
Qy	481	GTAATCATGTCGAACCTGATCGGCGGAAGACGGCTTGAACCACTAAACCCAGGGGCGG	540		
Db	1230	GTCCTTCACGTTCGACCCCGCGCAAGGCGAGAGGGGCTTCG---AATCGTTCGGGCCCC	1286		

```

QY 541 ATTGTCGCAATGCGCTGATCGTCGGGGTTTCCACTTCCCAATTTGGCCTTAGATGC 600
DB 1287 ATTGTGGCCAAAGCGCGATCGTGTGGCGGTTTCGACCTGCCAATATCGCCCTTGGCTGT 1346
QY 601 TTATATCTCGAGGACGATTCGCGACAGGGGTGAGAGCTGTGGCGCAACCTTTATCCG 660
DB 1347 TTTGTGTGCGGCGCATGATGAGACGCGGCGAAGAACTGTGGCGCAACTTATCTTCACTCCG 1406
QY 661 CAGCGCGCGAAGAGGGTGAAGAGACTTGGGGCAATGATTTGAGGCGCGCTGATGACC 720
DB 1407 CAGAGAGGTAAAGGGTGAAGAGAACTGGGGCAATGATTAAGAAACCCGCTGATGACC 1466
QY 721 GGGGTCTGGGGTCAATCACTATGATCCGCTGACGAACTGTGTCTTATGCTGACG 780
DB 1467 GGGGTCTGGGGCCAGATCAGTATGACCCCACTATTTGGTATTTTACGATGCTG 1526
QY 781 GGGGTGGGCCCAAGCTCCGAAACCAAGCGCGCAAGCGCGGCAAGCTGTATGCGACC 840
DB 1527 GCGGTAGGCGCGGATCCGAGGTTCAAGCGCGCAACCCGCGGCAAGCTTTATCGGCACC 1586
QY 841 AAGACCCGCTTTGCGGTGCTCCCGACACGCGGCAATTTCTGGGCTCAGCAGACCTTG 900
DB 1587 AAGACCCGCTTTGCTGCTCCCGACACGCGGCAAGTGTCTGGGCTCAGCAGACCTTG 1646
QY 901 CCGCGCGCAACTGGGACCAAGATGCACTTGCAGATGATGGTCCCAACGTCGATGTG 960
DB 1647 CCGCGCGCAACTGGGACCAAGATGCACTTGCAGATGATGGTCCCAACGTCGATGTG 1706
QY 961 CAACCTTCGCGCGAGATGAGGGTCTGCGCGCATCAACCCCAATGGCGGACGCGCGAG 1020
DB 1707 CAGCGCGCTCCGACATGAGAGCGGTGCAAGCCATCAACCCCAATGGCGGACGCGAG 1766
QY 1021 CCGCGTGTGCTGACGGGTGCGCTTGGCAAGACCGGACAGATGAGTGTGTTATGCGGCGC 1080
DB 1767 CGTGCGCTTTGACCGCGCTTCCGTGCAAAACCGGTACATGTGTGAGTTGACGCTGAA 1826
QY 1081 TCGGGCGAATTCGTGTGGGCGCGTGTATACCACTAGACCAATATGATGCGCTGATGAC 1140
DB 1827 ACGGGCGAATTCGTGTGGGCGCGGACACCACTAGACCAATATGATGATGATGAC 1886
QY 1141 GAGACGCGCTTTGAGACGCTGAGAACGAGATGCGGTGCTGAAAGAGCTGAGAGCTGAAAT 1200
DB 1887 GAAACCGGTGTGTCAGGTGAAATGAGATATCATCTTAAAGATGTGACACCGACTAC 1946
QY 1201 GAGCTGCGCGGACCTTCTGGGTGGGCGGACGATGCTGACGCGCACTGAAACCGGAC 1260
DB 1947 GCGATTTGCCGCAATTTCTTGGGTGAGCGGACTGGGCGGTGCGCATCTTGAACCCGAT 2006
QY 1261 ACCGGCATTTACTTCTTGGCGCTGAAACATGCTGCTAGATATTAATGCGCTGATCAA 1320
DB 2007 AGGGGCAATCTACTTCAATTCCTGAAACAGCCTGTGCGATTTGGCGGACGTGATCAA 2066
QY 1321 GAGTTTACGCGCTGACGCTCTATTAACACAGCGCGCAACCAAACTCGCGCGCGCTTT 1380
DB 2067 GAGTTTACGCGCATGAGAGCTCTAACACAGCGCGCACTTACTGTTGGCGCGGAAAAA 2126
QY 1381 GAAATATGAGCGCGCATGAGCGGATTAATGATGAGACGCGGCGCACTTGTGTGTGCGG 1440
DB 2127 GAAATATGAGCGCGCATGAGCGGATTAATGATGAGACGCGGCGCAACCTGTGTGTGCGG 2186
QY 1441 GAGGCGCTTGGCGGAACTATGCTGCGCGCTTTGTGACAGCGGCGGTTGTGTGCAAC 1500
DB 2187 GAGGCGCTTGGCGGAACTATGCTGCGCGCTTTGTGACAGCGGCGGTTGTGTGCAAC 2246
QY 1501 GGGCGGACGACCGCTATTTCCGTGCGCTTACGCGAGAAACCGGCGAGACTTTGTGTGCA 1560
DB 2247 GGGCGGACGACCGCTATTTCCGTGCGCTTACGCGAGAAACCGGCGAGACTTTGTGTGCA 2306
QY 1561 GCGCGTTCGAGAGCTGCGGACGCGGAGCGGACATCACTGATGAGAGCGGCGTGA 1620
DB 2307 ACCGCTGTGCGACTGTGCGCAGCGGTCAAGCATCACTGAGAGCGGCGTGA 2366

```

```

QY 1621 TATATGCGCATGCGTGGCGGCGGTCTGACTTATGGACAGCATTTGAACGCGCGCTGCC 1680
DB 2367 TATGTGGCATGCGAGGGGCGGTATATCTTACGCGACTTACCTGAAACAGCAATATCGGC 2426
QY 1681 GAGGCAATGATTCGACCTCGGTGCGGTATGCGATCTATGTCTTTGCACTGCGCAGTAA 1740
DB 2427 GCGACCATGATTCGACTTCAATGCGCAACGCGGTCTTACGCTTTCGCGCTTCCGCAATA 2486

RESULT 9
AAV29052
ID AAV29052 standard; cDNA; 1740 BP.
XX
AC AAV29052;
XX
DT 21-AUG-1998 (first entry)
XX
DE Alcohol and/or aldehyde dehydrogenase A' nucleotide sequence.
XX
KW Alcohol/aldehyde dehydrogenase A' enzyme; recombinant organism; aldehyde;
KW ketone; carboxylic acid; L-sorbose; D-sorbitol; 2-keto-L-gulonic acid;
KW L-ascorbic; inhibition; ds.
XX
OS Gluconobacter oxydans.
XX
FH Key
FT 1. 1740
FT CDS
FT
FT /tag= a
FT /product= "alcohol and/or aldehyde dehydrogenase A'"
XX
PN EP832974-A2.
XX
PD 01-APR-1998.
XX
PT 11-SEP-1997; 97EP-00115801.
XX
PR 19-SEP-1996; 96EP-00115001.
XX
PA (HOFF ) HOFFMANN JA ROCHE & CO AG F.
XX
PI Asakura A, Hoshino T, Ojima S, Shinjoh M, Tomiyama N;
XX WPI: 1998-195228/18.
XX DR P-PSDB; AAM37874.
XX
PT Recombinant Gluconobacter oxydans alcohol and/or aldehyde dehydrogenase
PT enzyme(s) - useful for converting substrate(s), e.g. L-sorbose or D-
PT sorbitol to 2-keto-L-gulonic acid.
XX
PS Disclosure: Page 29-30; 59pp; English.
XX
XX
CC This is the nucleotide sequence encoding the Gluconobacter oxydans
CC alcohol and/or aldehyde dehydrogenase A' enzyme. The enzymes or
CC recombinant organisms can be used to convert suitable substrates to
CC aldehydes, ketones or carboxylic acids, especially to convert L-sorbose
CC or D-sorbitol to 2-keto-L-gulonic acid, which can be converted to L-
CC ascorbic acid by standard procedures. The derivatives of ADH enzymes
CC have desited substrate specificity, higher affinity to a substrate, lower
CC affinity to an inhibitory compound, higher stability against temperature
CC and/or pH and higher catalytic speed
XX
SQ Sequence 1740 BP; 347 A; 576 C; 503 G; 314 T; 0 U; 0 Other;

Query Match 60.6%; Score 1053.6; DB 2; Length 1740;
Best Local Similarity 75.7%; Pred. No. 2.2e-254;
Matches 1318; Conservative 0; Mismatches 419; Indels 3; Gaps 1;

QY 1 ATGAACCCCAACAGCTGCTTGCACAGCGGCGGTATGCTTACCGCGCCCGCC 60
DB 1 ATGAAGAGCTGCTTCTTCTGCTGAGCGCTTGCAGAGCTTGAAGCTATAGCTCTTT 60
QY 61 GCATTGCGGAGGTAAACCCGATTAACGATGCTGTGGGAAACCCGCGCTGTGA 120

```

Db 61 GGCGTTGCTCAAGTGAACCCCGGTCACCGATGAAATGCTGGCGAACC CGCGCTGGTGA 120  
 QY 121 TGAATTAACAGCCCGCAACCAAGAAATATTCGCCACTCGCCCTGACCCAGATAC 180  
 Db 121 TGAATTAACAGCCCGCAACCAAGAAATATTCGCCACTCGCCCTGACCCAGATAC 180  
 QY 181 GCCGACAAAGTGTGATGTAAGTGTGAGGCGCGGAGTGAAGCGGAGCGCTA 240  
 Db 181 ACTGAGAAAGCTCGGCCAATGCACTGCTGTGGCGCGGCGCATGAGCGGAGCAAAATC 240  
 QY 241 CAGGTCAACGCCGATGATCCATGATGAGCGGTATATTCGCAAAACCCCGGTGATGATC 300  
 Db 241 CAGGTCAACGCCCGCTGATCCATGATGAGCGGTATATTCGCAAAACCCCGGTGATGATC 300  
 QY 301 CAGGCGGTGATGCGCAAAACAGGCGATTCGATTCGGGAAACACCGCGGCAATCGCCGCG 360  
 Db 301 CAGGCGCATGACCGCAAAACAGGCGATTCGATTCGGGAAACACCGCGGCAATCGCCGAGC 360  
 QY 361 GTGCGCAACGCTAAACCGCCAAAGCGACCGCAAGCGCGCGCTTCCTTACGCGACAGAC 420  
 Db 361 ATGCGCAACGCTGAAACAGCTTTGGCGAGCGCACCGCGCGCATGGCGCTGTACGCGACCAAC 420  
 QY 421 CTGATTTTCAGCTCAGGGCAACACATTCGAGCGCTGATATGAGAGCGGCGACAGTC 480  
 Db 421 GTTACTTGTGTTCTGAGCAACACCTGCTGCTGACACCGCAACTGACCAATG 480  
 QY 481 GTATTCGATGCGAAGCGTGAATCGGCGAAGACGCGCTTGAACGATTAACACACGCGGCGCG 540  
 Db 481 AGCTTGACGCTGACACCGCGCGCAAGCGAAGACATGATTTCCG--AACTGTCGCGGCGCG 537  
 QY 541 ATTGTGCGCAATGCGCTGATTCGTCGCGGCTTCACCTGCGCAATTCGCGCTATGATGAC 600  
 Db 541 ATGTGTGCAAAACGCGCGTGAATGTTGCGGCTTCGACCTGCGCAATTCGCGCTATGATGAC 597  
 QY 601 TTATCTCGGAGGACGATTCGCGAGCGGAGTGAAGATGAGCGCAACCACTTTATCCCG 660  
 Db 598 TTTGTCTCGGCGCAACGACTGGCCACCGGTGAAGACTGTGGCGCAACTTCAATCCCG 657  
 QY 661 CAGCGCGGCGAAGGAGTGAAGACTTGGGCGCAATGATTTGAGGCGCGCTGATGAC 720  
 Db 658 CCGGCTGCGCAAGAGGTGATGAGACTTGGGCGCAAGATTAACCAACCGGTTGAGTGAAC 717  
 QY 721 GGGGTGCGGAGTGAAGTCACTATGATCCGTCAGCAACCTTGTGTTATGAGCTGAC 780  
 Db 718 GGGGTGCGGAGTGAAGTCACTATGATCCGTCAGCAACCTTGTGTTATGAGCTGAC 777  
 QY 781 GGGGTGCGGAGTGAAGTCACTATGATCCGTCAGCAACCTTGTGTTATGAGCTGAC 840  
 Db 778 GGTGTTGCGCGGCTTGGAAACCCAGCGCGCACACCGCGCGCACTGTAACGCGCAC 837  
 QY 841 AAGACCGGCTTGGCGGTGCGTCCGACACGCGGAGATGTTGCGGTGACCAAGACCTG 900  
 Db 838 AAGACCGGCTTGGCGGTGCGTCCGAGACTGCGAGATGCTGCGGTGACCAAGCTG 897  
 QY 901 CCGCGGACAACTGGAGCAACAGATGACGTTGAGATGATGGTGCACCACTGATGATG 960  
 Db 898 CCGCGGACAACTGGAGCAACAGATGACGTTGAGATGATGGTGCACCACTGATGATG 957  
 QY 961 CAACCTTCGCGCGAGATGAGGCTTCGCGGCACTCAACCCCAATCGCGGACGCGGAG 1020  
 Db 958 CAGCGCGGACGATGACAGCGGCTGCGTCCGATCAACCCGACACCGCGACGCGGAG 1017  
 QY 1021 CGCGGTGCTGAACGCGTGCCTTGGCAAGACGCGACGATGATGCTGTTGATGCGGCG 1080  
 Db 1018 CGTGCCTTTCGACCGCGCTTCGCTGCAAAACCGGACCAATGGAGTTGACGCGGAA 1077  
 QY 1081 TCGGCGCAATTCCTGTGCGCGCTGATACCACTAATGATTCGCTTCGATGAC 1140  
 Db 1078 ACCGCGCAATTCCTGTGCGCGCTGATACCACTAATGATTCGATGATGAC 1137  
 QY 1141 GAGACCGGCTTGAACGATGAGATGCGGTGCGTGAAGAGCTGAGCTTGAAT 1200  
 Db 1138 GAAAAAGGCACTGATGACCGTGCAGAGTGCAGAAAGTTCTGACGAGTGAACCCCTAT 1197

QY 1201 GAGCTGCGCGACCTTCCTGAGGAGCGGACGAGTGTGTCAGCGCATGAACCGGAC 1260  
 Db 1198 GAGCTGCGCGCGGCTGCTGAGTGTGAGCGTGAATGAGCGCGTGTGAGACCGGAT 1257  
 QY 1261 ACCGCGATTTAATTCTTCCGCTGAGCAATGCTGCTGCAATTTATGCGCTTGAATCA 1320  
 Db 1258 ACCGCGATTTAATTCTTCCGCTGAGCAATGCTGCTGCAATTTATGCGCTTGAATCA 1317  
 QY 1321 GAGTTTACGCGCTGAGAGTCTAATAACAGGCGGACCGCAAAATTCGCGCGGCTTT 1380  
 Db 1318 GAATTCAGCTGCTGAGAGTCTAATAACAGGCGGACCGCAAAATTCGCGCGGCTTT 1377  
 QY 1381 GAAATATGAGCGCGCATGACGCGATGATGATGACACCGGCGCATCTTGTGTGCGCG 1440  
 Db 1378 GAGCTGTTGGCGGTATGACGCGCATGACATGACGACAGCGCGCATCCCTGTGAGCGCT 1437  
 QY 1441 GAGCGCGCTGCGGGAATCTGCGCGCTTTGTGCAACGCGGAGCGGTGTGATCAAC 1500  
 Db 1438 GAGCGGAAAGCTGAACTAGCGGCGCTGTCTGTGACCGGCTGCGGCGCTTGTTCAC 1497  
 QY 1501 GCGCGGACCGACGCTATTTCCGTGCGCTGACGCGAAGAAACCGCGAGACTTTGTGCGAG 1560  
 Db 1498 GCGCGGACCGACGCTATTTCCGTGCGCTGACGCGAAGAAACCGCGAGACTTTGTGCGAG 1557  
 QY 1561 GCGCGCTTCTGAGAGGTCGAGCGGCGAGCGGATGACGATGACGATGAGCGGCGTGC 1620  
 Db 1558 ACCGCTTGGGAGTGTGCGCTTGGGCGCAAGCTGTCTGACGAGATGACGCGGCTCAA 1617  
 QY 1621 TATATCGCATCGGTGCGGCGGCTGACCTATGACGCAATGGAACGCGCGCTGCGC 1680  
 Db 1618 TACATCGCATCGGCGGCGGCGGCGGAGCACTATGATGTTGTTCCAAACCGTCCCTGAGC 1677  
 QY 1681 GAGCGAATCGATTTGACCTGCGTGTGATGCGATCTATGCTTTTGTGATCGCGGACTA 1740  
 Db 1678 GAGCGGTCGACTGACCGGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1737

RESULT 10  
 ADI34116  
 ID ADI34116 standard; DNA, 1740 BP.  
 XX  
 AC ADI34116;  
 XX  
 DT 15-APR-2004 (first entry)  
 XX  
 DE Ketogulononicigenium sp. sorbitol dehydrogenase (SDH) 2 DNA.  
 XX  
 XX Sorbitol dehydrogenase; SDH; cytochrome c; molecular biology;  
 KW bacteriology; industrial fermentation; L-sorbitose; 2-keto-L-gulononic acid;  
 KW 2KLG; gene; ds.  
 XX  
 CS Ketogulononicigenium.  
 XX  
 FH Key  
 FT CDS Location/Qualifiers  
 FT /\*tag= b  
 FT /product= "SDH protein"  
 FT sig\_peptide  
 FT /\*tag= a  
 FT /\*tag= 70..1737  
 FT /\*tag= c  
 FT /product= "Mature SDH protein"  
 XX  
 PN US200328672-A1.  
 XX  
 PD 11-DEC-2003.  
 XX  
 PF 06-JUN-2002; 2002US-00162713.  
 XX  
 PR 06-JUN-2002; 2002US-00162713.  
 XX  
 PA (ARCH ) ARCHER-DANIELS MIDLAND CO.

XX Choi E, D'elia J, Kim H, Kim M, Lee JK, Pan J, Stoddard SF,  
PI Yun D;  
XX WPI; 2004-052025/05.  
DR P-PSDB; ADI34120.  
XX  
PT New proteins of sorbitol dehydrogenases and cytochrome c of the strains  
PT ketogluligenium spp., useful in molecular biology, bacteriology and  
PT industrial fermentation, e.g. for producing L-sorbose and 2-keto-L-  
PT gulonic acid.  
PS Claim 21; SEQ ID NO 3; 68bp; English.  
XX  
CC The invention relates to the identification and isolation of nucleic acid  
CC sequences and proteins of sorbitol dehydrogenases (SDH) and cytochrome c  
CC of the strains, Ketogluligenium sp. The proteins and nucleic acid  
CC molecules are useful in the fields of molecular biology, bacteriology and  
CC industrial fermentation specifically for producing L-sorbose and 2-keto-L-  
CC -gulonic acid (2KLG). The present sequence is Ketogluligenium sp. SDH  
CC DNA.  
XX  
SQ Sequence 1740 BP; 369 A; 567 C; 494 G; 310 T; 0 U; 0 Other;  
Query Match 60.5%; Score 1052; DB 12; Length 1740;  
Best Local Similarity 75.7%; Pred. No. 5.ee-254;  
Matches 1317; Conservative 0; Mismatches 420; Indels 3; Gaps 1;  
QY 1 ATGACCCCAACAGCTGCTTCCGACCGCGCGCTGCTATTGCTTACCGGCGCCG 60  
Db 1 ATGAGACGAGAGCTTTTCTGTTGACGCGTGTCTGCGCTTGAAGCTACCGCAATTT 60  
QY 61 GCATTCCGCGAGGTAACCCGATTACCGATGACTGCTGCGAACCAGCCGCTGCTGTA 120  
Db 61 GCGCTTGCTATGTGACCCCGCTCACCGAGACTGTGTGCAAAACCCGCGCGCGAA 120  
QY 121 TGGATTACTACGCGCGCAACCAAGAAATCTATGCGCACTCGCCCTGACCCAGATCACT 180  
Db 121 TGGATCAGCTACGCGCGCAACCAAGAAATCTATGCGCACTCGCCCTGACCAAGATCAAG 180  
QY 181 GCGGACAAAGTGTGCTGCACTGTTGCACTGCTGCGCGCGGATGAGAGCGGCGCGTA 240  
Db 181 CCGGAGAACTCGTGTGCACTGCTGCACTGCTGCGCGCGGATGAGAGCGGCGCGTAAGTC 240  
QY 241 CAGGTCAAGCGATGATCATGATGCGGTGATCTGTGCAAAACCCCGGTATGTGATC 300  
Db 241 CAGGTCACTCGCTGATCCATGATGCGGTGATGTACTGTGCGAAACCCCGCGCATCATC 300  
QY 301 CAGGCGCTGATGCGCAACAGCGCATCTGATCTGGAACAACCGCGCGCACTGCGCGCC 360  
Db 301 CAGGCGATCGACGTAAACCGCGCATCTGATCTGGAACAACCGCGCGCACTGCGCGCAAC 360  
QY 361 GTGCGCAAGCTAAACCGCGCAAGCGCGCAAGCGCGCGCTGCGCTTTACGCGCAAGC 420  
Db 361 GTGCGCAAGCTAAACCGCGCTTGTGAGCGCATCGCGGTATGCGGTGAGCGCAACAC 420  
QY 421 CTCTATTGAGCTCATGAGCAACCATCTGATCGCGCTGATATGAGACGCGCGCATGCTC 480  
Db 421 GTTTACTTCTCTGCGGGAACAACCTGTGTGCTGTGACGCAACCGCGCGCAAGTC 480  
QY 481 GTATTGATGTGCAAGTGTGAGCGGCGAAGACGCGCTTGACAGTAAACCAACGCGCGCG 540  
Db 481 AGCTTGACGTGACCGCGCGCAAGCGAAGACATGCTTTC--TAACTGTGTGCGCGCG 537  
QY 541 ATTGTGCGCAATGCGCTCATGTGTGCGGGTTCACTGCGCAATATTTGCGCTTATGATGC 600  
Db 538 ATGTGCTTAACGCGCGTATCTGTGCGCGGTTCGACCTGCGCAATATCTGCGCTTCTGCTGC 597  
QY 601 TTATATCTGCGGCGACGATTCGCGACGCGGTGAGAGACTGTGCGCGCAACCACTTTATCCG 660  
Db 598 TTGCTTTGCGGCGCATGCGGACTACCGGTGAGAGAACTGTGCGCGCACTTCTATCTCC 657  
QY 661 CAGCCGCGCGAAGAGGTGACGAGACTTGGGCAATGATTTGAGGCGCGCTGATGACC 720

Db 658 AAGCGGCTGAAGAGGAGATGAACCTGTGGGCAACGACTACGAACCCGCTGATGACC 717  
QY 721 GCGCTGTGGGATCAGATCACTATGATCCCGTGAACGAACTTGTGTTATGCTGACAC 780  
Db 718 GCGCTGTGGGCAAAATACGATACGACCCCGTCAACCACTGTATTTACGATCGTCG 777  
QY 781 GCGGTGGGCGGACGCTCCGAAACCCAGCGCGGACACCGCGCGGACGCTGTATGACAC 840  
Db 778 GCGGTGGGCGGCTTTCGAAACCCAGCGCGGACACCGCGCGGACGCTGTATGACACG 837  
QY 841 AACACCGCTTGTGCGGTGCGGTCCGACACGCGCGGATGTCTGCGCTCAACAGACCTG 900  
Db 838 AACACCGCTTGTGCGGTGCGGTCCGACACCGCGCGGATGTCTGCGCTCAACAACTG 897  
QY 901 CCGCGCGCAACTGGGACCAAGATGCAAGTTGAGATGATGTGCGCAACGCTGATGTG 960  
Db 898 CCGCGCGCAACTGGGACCAAGATGCAAGTTGAGATGATGTGCGCAACGCTGATGTG 957  
QY 961 CAACCTGCGCGGATGAGGAGGTGCGCGGCAATCAACCCCAATGCGCGGACGCGGAG 1020  
Db 958 CAGCTTGTGCGGTGATGAGGAGGTGCGCGGCAATCAACCCCAATGCGCGGAG 1017  
QY 1021 CCGCGGTGTGACGCGGTGCGCTTGAAGACCGCGCAAGATGTGCTGTTGATGCGGCG 1080  
Db 1018 CGTGTGTGTGACGCGGTGCGCTTGAAGACCGCGCAAGATGTGCTGTTGATGCGGCG 1077  
QY 1081 TCGGCGCAATTCCTGTGGGCGCGGTGATGCAATCAACCAATATGATGCTGATGAC 1140  
Db 1078 ACGGCGCAATTCCTGTGGGCGCGGTGATGCAATCAACCAATATGATGCTGATGAC 1137  
QY 1141 GAGACCGGCTTGTGACGCGGTGAGAGAGATGCGGTGCAAGAGCTGCAAGCTGATAT 1200  
Db 1138 GAAACCGGCTTGTGACGCGGTGAGAGAGATGCGGTGCAAGAGCTGCAAGCTGATAT 1197  
QY 1201 GAGCTGTGCGGACCTTCTGTGGGCGCGGTGATGCAATCAACCAATATGATGCTGATGAC 1260  
Db 1198 GCGATTTGCGGCAATTCCTGTGGGCGCGGTGATGCAATCAACCAATATGATGCTGATGAC 1257  
QY 1261 ACCGCACTTCTCTTCCGCTGCAACCAATGCTGCTGATATGATGCTGATGATGAT 1320  
Db 1258 AGCGGCACTTCTCTTCCGCTGCAACCAATGCTGCTGATATGATGCTGATGATGAT 1317  
QY 1321 GAGTTAGCGGCTGACGCTTATPAACACAGCGCGACCGCAAACTTGTGCGCGGCTTT 1380  
Db 1318 GAGTTAGCGGCTTATPAACACAGCGCGACCGCAAACTTGTGCGCGGCTTT 1377  
QY 1381 GAAATATGCGCGCGCATGAGCGGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
Db 1378 GAAATATGCGCGCGCATGAGCGGATGATGATGATGATGATGATGATGATGATGATGAT 1437  
QY 1441 GAGCGCGCTGCGGGAATCTCTGCGCGGCTTGTGCAACGCGCGGCGGTGTGTTCAAC 1500  
Db 1438 GAGCGCGCTGCGGGAATCTCTGCGCGGCTTGTGCAACGCGCGGCGGTGTGTTCAAC 1497  
QY 1501 GCGGGAACGACGCTATTTTCCGTGCTTCAACGCGGCAACCGCGGACCTTTGTGCGAG 1560  
Db 1498 GCGGGAACGACGCTATTTTCCGTGCTTCAACGCGGCAACCGCGGACCTTTGTGCGAG 1557  
QY 1561 GCGGCTTGTGCGAGCGGTGCGACAGGCGGCGGATGATGATGATGATGATGATGATGAT 1620  
Db 1558 ACCGCTGTGCGGACGCTGCTTCCGCGCAAGCGGCTGTGTGATGATGATGATGATGAT 1617  
QY 1621 TATATGCGATCGGTGCGGCGGTGCTTCACTATGCGGCAATGAAACGCGCGCTGCGC 1680  
Db 1618 TATATGCGATCGGTGCGGCGGTGCTTCACTATGCGGCAATGAAACGCGCGCTGCGC 1677  
QY 1681 GAGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740  
Db 1678 GAGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1737

RESULT 11

AD134123  
ID AD134123 standard; DNA; 3200 BP.

AC AD134123;

DT 15-APR-2004 (first entry)

XX Ketogulonigenium sp. gene encoding cytochrome c and SDH proteins.

XX Sorbitol dehydrogenase; SDH; cytochrome c; molecular biology;

KM bacteriology; industrial fermentation; L-sorbose; 2-keto-L-gulonic acid;

XX 2KLG; gene; ds.

OS Ketogulonigenium.

XX Location/Qualifiers

FT Key 652..655

FT RBS /\*tag= a

FT CDS /standard\_name= "Shine-Dalgarno"

FT sig\_peptide 663..1157

FT mat\_peptide 663..764

FT RBS /\*tag= c

FT sig\_peptide 663..764

FT mat\_peptide 663..764

FT RBS /\*tag= d

FT sig\_peptide 663..764

FT mat\_peptide 663..764

FT RBS /\*tag= e

FT sig\_peptide 663..764

FT mat\_peptide 663..764

FT RBS /\*tag= f

FT sig\_peptide 663..764

FT mat\_peptide 663..764

FT RBS /\*tag= g

FT sig\_peptide 663..764

FT mat\_peptide 663..764

FT RBS /\*tag= h

FT sig\_peptide 663..764

FT mat\_peptide 663..764

FT RBS /\*tag= i

FT sig\_peptide 663..764

FT mat\_peptide 663..764

FT RBS /\*tag= j

FT sig\_peptide 663..764

FT mat\_peptide 663..764

FT RBS /\*tag= k

Best Local Similarity 75.7%; Pred. No. 6.6e-254; Matches 1317; Conservative 0; Mismatches 420; Indels 3; Gaps 1;

QY	1	ATGAACCCCAACAGCGCTTCGACACGAGCGCGCGCTGATGCTACCGCGCCGCC	60
DB	1241	ATGAAGACGAAGTCTTTCTGTTGTCAGCGCTTCTGCTGCAAGCTACGACCAATT	1300
QY	61	GCATTGCGGAGGTAACCCGATTAACGATGAATGCTGCGGAACCCGCGCTGTGAA	120
DB	1301	GCCTTGTGATGACCCCGTCACCGACGACTGTGCAAAACCCGCGCGCGAA	1360
QY	121	TGATTAACATGAGCGCGCAACCAAGAAATATATGCGACATGCGCGCTGACCAT	180
DB	1361	TGATGAGCTACGCGCGCAACCAAGAAATATATGCGACATGCGCGCTGACCAT	1420
QY	181	GCCGACACGTTGCTGATGCAATGCTGCTGCGCGCGCGGATGAGCGCGCGTA	240
DB	1421	CCGAGAACGTCGCTGCTGCTGCAATGCTGCTGCGCGCGCGGATGAGCGCGTA	1480
QY	241	CAGGTACGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	300
DB	1481	CAAGTCACTCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG	1540
QY	301	CAGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	360
DB	1541	CAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1600
QY	361	GTCGCCACGTTAAAGCCCAAGCGGACCGGACCGGCGCTGCTTACGCGACGAG	420
DB	1601	GTCGCCACGTTAAAGCCCAAGCGGACCGGACCGGCGCTGCTTACGCGACGAG	1660
QY	421	CTCATTTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG	480
DB	1661	GTTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1720
QY	481	GATTCGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG	540
DB	1721	ACGTTGACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG	1777
QY	541	ATTGTCGCAATGCTGATGATGATGATGATGATGATGATGATGATGATGATG	600
DB	1778	ATGTCGCTAAAGCGGATGATGATGATGATGATGATGATGATGATGATGATG	1837
QY	601	TTATCTCGGCGCAAGTTCGCGACCGGATGATGATGATGATGATGATGATGATG	660
DB	1838	TTGCTTCGCGCGCAATGATGATGATGATGATGATGATGATGATGATGATGATG	1897
QY	661	CAGCGCGCAAGGAGGATGATGATGATGATGATGATGATGATGATGATGATG	720
DB	1898	AAAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1957
QY	721	GAGCTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	780
DB	1958	GAGCTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2017
QY	781	GAGCTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	840
DB	2018	GAGCTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2077
QY	841	AAACCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	900
DB	2078	AAACCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2137
QY	901	CCGCGCGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG	960
DB	2138	CCGCGCGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2197
QY	961	CAACCTTGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATG	1020
DB	2198	CAGCTTGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATG	2257
QY	1021	CGCGCTGCTGACGCGGATGATGATGATGATGATGATGATGATGATGATGATG	1080

Query Match 60.5%; Score 1052; DB 12; Length 3200;

Db 2258 CGTGGCTGCTGACCGCGGCTTCCGTGCAAAACCGGTACCATGTGGAGATTGACGCTGAA 2317  
 Qy 1081 TCGGGCGAATTCCGTGGGCGCGGTGATACCACTACCAATGATGATGCGCTGATGAC 1140  
 Db 2318 ACGGGCAATTCCTGTGGGCGCGGACACCACTACCAATGATGATGATGATGAC 2377  
 Qy 1141 GAGACCGGCTTGTGACGCTGAAACGAGATGCGGTGTAAGAGCTGACGTTGATAT 1200  
 Db 2378 GAAACCGGCTGTGCTGACGCTGAAAGATATCATCTAAAGATCTGACACCGACTAC 2437  
 Qy 1201 GAGCTGCGCGCACTTCCGTGGGCGCGGACGCTGCTGACGCGACCTGAAACCGGAC 1260  
 Db 2438 CGCATTTGGCCGATTTCTTGGGTGAGCGACCTGCGCGGTGCGATCTTGAACCCCGAT 2497  
 Qy 1261 ACCGCGATTTACTTCTTGGCGCTGAAACATGCTGCTACGATATTAATGCGCTTGAATCA 1320  
 Db 2498 AGCGGCACTTACTTCAATTCCTCCCTGAAACAAGCGCTGTGGGAGTGCATCA 2557  
 Qy 1321 GAGTTTGGCGGCTCGACGCTTATTAACACGAGCGGACCGCAAACTCGCGCGGCTTT 1380  
 Db 2558 GAGTTCAAGCAATGACGCTTACACACGAGCGGCTTACCTGCTTGGCGGAAAA 2617  
 Qy 1381 GAAATATGGGCGCGCATGACGCGATGATATCAGCACCGGCGGCACTTGTGCTGGCG 1440  
 Db 2618 GAAATATGGGCGCGCATGACGCGATGATATCAGCACCGGCAAAACCTGTGTGCTGTC 2677  
 Qy 1441 GAGCGCCCTGGCGGAACTACTGCGCCGTTTGTGACGCGGCGGCTGTGTGTTCAAC 1500  
 Db 2678 GAGCGTGTGGCGTGAATCTACTGCGCGCTCTCGACGCGGCTGTGCTGTTCAAC 2737  
 Qy 1501 GCGGCGGCGGCGGCTTTCCTGCGCTGAGCGGAAACCGGCGGAGATTTGTGCGAG 1560  
 Db 2738 GCGGCGGCGGCTTTCCTGCGCTGAGCGGAAACCGGCGGAGATTTGTGCGAG 2797  
 Qy 1561 GCGCGTTCGACGCTGCGGACGCGGCGGCGGCGGCTGACGATGATGAGCGGCTGCA 1620  
 Db 2798 ACCGCTGCGGCTGCTGCGCTGCGGCGGCGGCGGCTGACGATGATGAGCGGCTGCA 2857  
 Qy 1621 TATATGCGGCTGCGTGGGCGGCTGCTGCTGATGAGCGGCTGATGAGCGGCTGCGC 1680  
 Db 2858 TATATGCGGCTGCGTGGGCGGCGGCGGCTGCTGCTGATGAGCGGCTGCGC 2917  
 Qy 1681 GAGGCAATCGATTCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740  
 Db 2918 GAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2977

## RESULT 12

ADE94132  
 ID ADE94132 standard; DNA; 1827 BP.

AC ADE94132;

DT 12-FEB-2004 (first entry)

DE Alcohol/aldehyde dehydrogenase coding sequence, SEQ ID 2.

XX Enzyme: alcohol/aldehyde dehydrogenase; pyrrolo quinoline quinone;  
 KM L-sorbose; L-sorbose; 2-keto-L-gulononic acid; gene; ds.

XX Pseudoglucobacter sacccharoketogenes; IF014464.

OS Pseudoglucobacter sacccharoketogenes; IF014464.

FH Key location/Qualifiers  
 FT 1.1827  
 FT CDS /product= "Alcohol/aldehyde dehydrogenase"

XX JF2003159079-A.

XX 03-JUN-2003.

XX 29-NOV-2001; 2001JP-00364508.

XX

PR 29-NOV-2001; 2001JP-00364508.  
 XX  
 PA (FUT) FUJISAMA PHARM CO LTD.  
 XX  
 DR WPI, 2003-818681/77.  
 DR F-PSDB; ADE94131.  
 XX  
 PT Novel alcohol/aldehyde dehydrogenase protein catalyzing oxidation of  
 PT hydroxymethyl group of compound to aldehyde group and aldehyde group of  
 PT compound to carboxyl group, useful for manufacturing 2-keto-L-gulononic  
 acid.  
 PS Claim 6; SEQ ID NO 2; 48bp; Japanese.  
 XX  
 CC The present invention relates to an alcohol/aldehyde dehydrogenase (I;  
 CC ADE94131), which catalyzes oxidation of hydroxymethyl group of a compound  
 CC to an aldehyde group and the aldehyde group of a compound to a carboxyl  
 CC group. (I) does not contain hemiterium and rare earth elements. (I) has  
 CC an optimum pH of 4.5-5.5, isoelectric point of 4.1 +/- 0.3, pyrrolo  
 CC quinoline quinone as a prosthetic group and K<sub>m</sub> value of 40 mM for  
 CC sorbose. (I) is useful for manufacturing a compound having a carboxyl  
 CC group by contacting the compound having a hydroxymethyl group or an  
 CC aldehyde group with (I) where the compound having a hydroxymethyl group  
 CC or an aldehyde group is L-sorbose or L-sorbose and the compound having  
 CC a carboxyl group is 2-keto-L-gulononic acid.  
 XX  
 SQ Sequence 1827 BP; 342 A; 612 C; 551 G; 322 T; 0 U; 0 Other;

Query Match 36.0%; Score 625.8; DB 10; Length 1827;  
 Best Local Similarity 61.8%; Pred. No. 6.2e-147;  
 Matches 1066; Conservative 0; Mismatches 647; Indels 12; Gaps 4;

Qy 24 CACGAGCGCGCGCTGCTATTGCTTACCGCGCGCGCGCATTCGCGAGTAACCCGAT 83  
 Db 99 CACCG 158  
 Qy 84 TACCGATGACGCTGCGGGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 143  
 Db 159 CAGGCTGACGATTCGCGCGGGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 218  
 Qy 144 AGAAATATGCG 203  
 Db 219 CAGGCGTGGGATTAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 278  
 Qy 204 ACTGCTGCGGCG 263  
 Db 279 GCTGCTGCG 338  
 Qy 264 TGGCGTATGATTCGCGGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 323  
 Db 339 CGGCGTATTCCTCGCGGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 398  
 Qy 324 GATCTGATTCGCGGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 380  
 Db 399 TCCCTCATTCGCGGAATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 458  
 Qy 381 AGGCGACGCGAAGCG 440  
 Db 459 CGGCG 518  
 Qy 441 CAACCATGATTCG 500  
 Db 519 CAATTCGCTTTCG 578  
 Qy 501 ATCGGCGGAAGCG 560  
 Db 579 CCGGCG 638  
 Qy 561 CGTGGCGGTTCACTGCGCAATATTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 620  
 Db 639 GATGCGAGGCTCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 698  
 Qy 621 CGGAGCGGTTAGGAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 680

[illegible]

Db	1776	GGCTTCGGGCGGCAAGGCATCTACGCTTCCTCGCTCCGAGAA	1820
		RESULT 13	
		AAQ13580	
		ID AAQ13580 standard; DNA; 2214 BP.	
		AC AAQ13580;	
XX	XX	27-AUG-2003 (revised)	
XX	XX	25-MAR-2003 (revised)	
XX	XX	09-DEC-1991 (first entry)	
XX	XX	A.alcoccigenes membrane-bound ADH 72kD sub-unit.	
XX	XX	alcohol dehydrogenase complex; carboxylic acid production; ss.	
XX	XX	Acetobacter sp.	
XX	XX	EP48969-A.	
XX	XX	02-OCT-1991.	
XX	XX	26-FEB-1991; 91EP-00102793.	
XX	XX	26-FEB-1990; 90UP-00042391.	
XX	XX	26-MAR-1990; 90UP-00073440.	
XX	XX	(NAKA-) NAKANO VINEGAR CO LTD.	
XX	XX	Tamaki T, Takekura H, Tayama K, Fukaya M, Okumura H, Kawamura Y;	
XX	XX	WPI; 1991-289462/40.	
XX	XX	P-PSDB; AAR13993.	
XX	XX	Gene for membrane-bound alcohol dehydrogenase complex - obtd. from	
XX	XX	Acetobacter alcoccigenes, used for prodn. of enzyme for converting	
XX	XX	alcohol to acid.	
XX	XX	Claim 3; Page 20-22 and Fig 3; 36pp; English.	
XX	XX	Total DNA was prepared from A.alcoccigenes NH-24, digested with PstI and	
XX	XX	SmaI and ligated to PstI-SmaI cleaved pUC18. The ligation mixture was	
XX	XX	used to transform E.coli JM109. Probes were designed based on the N-	
XX	XX	terminal amino acid sequence of the ADH complex isolated from A.	
XX	XX	alcoacetigenes (see AAQ13582-Q13584). The gene encoding the 72kD ADH	
XX	XX	subunit was isolated and sequenced. It forms a membrane-bound ADH complex	
XX	XX	immediately downstream of the gene encoding the 72kD sub-unit. See also	
XX	XX	AAQ1381 (updated on 25-MAR-2003 to correct PA field.) (Updated on 27-	
XX	XX	AUG-2003 to correct OS field.)	
XX	XX	Sequence 2214 BP; 451 A; 683 C; 662 G; 418 T; 0 U; 0 Other;	
XX	XX	Query Match 5.1%; Score 88.8; DB 2; Length 2214;	
XX	XX	Best Local Similarity 50.8%; Pred. No. 4,9e-12;	
XX	XX	Matches 266; Conservative 0; Mismatches 252; Indels 6; Gaps 2;	
Qy	52	GGCCCCCGCCGACATTGCCGCAAGTAACCCCATTCACGATTAACCTGCTGGGAAACCGGCC	111
Db	97	GCAATCGCAGATGACGGGCGAGGCGCCACGGGGAGAACGATCAATCAATGCGATGATCAC	156
Qy	112	GCTGTGGAATGATTAACCTACGGCCGCAACAGAAAGAACTATGCGCACTGCCCTGACC	171
Db	157	CCCGGTAATGATGATGACTATGCGCGACACATTCATGACAGGGCTACAGCCCGTGAAT	216
Qy	172	CAGATCACTGCCGACACAGTTGTCAGTTGCATCTGTCCTGCGGCCCGGAGATGAGGGC	231
Db	217	CAGATCAACCGTTCCAAATGTCGTAACTGTAAAGCTGGCCCTGTATCTGACCTTGATACC	276
Qy	232	GGGGCGGTACAGT---CACGCCGATGATCATATGAGCTGATGTATGCGCAACCCC	288
Db	277	AACCGTGGCCAGGAAGCAGCCGCTCTGTTATTTATGAGCCTATGTAAGCAGCAACCAAC	336



QY 289 GGTGATGATGATCCAGGCGGTGATGCGGCAACAGGCGATGCTGATCTGGAAACACCCGCC 348  
 DB 337 TGGAGCATGATGAAAGCGTGCAGCGCGCAACCGGCAAGCTGCTGTGCTTATGACCG 396  
 QY 349 CAATGCGCCGCGTCCGCAAGCTTAAACGCCCAAGGCGAC---CGCAAGCGCGCGTCCG 405  
 DB 397 CGCGTCCCGCAACATGTCGCAAGAGGCTGCTGTGACACGGTCAACCGTGGCGGCA 456  
 QY 406 CTTTACGGCACAGCCCTTATTTCACTCATGGGAACAACATGATCGCGCTGATATG 465  
 DB 457 TACTGGAATGGCAAGGCTATTTTCGCAAGTTCGACGGTGCCTGATTCGGCTGACGCC 516  
 QY 466 GAGACGGGCGGAGTTCGATTCGATTCGACGATGCGGCGGAAGACGGCTTGACCAATG 525  
 DB 517 AAGACCGGCAAGCTGTGCTGAGCGTCAACCAATTCCGCCGGAAGCGGAACCTGGGCAAG 576  
 QY 526 AACACCAAGGCGCGGATTTGCGCAATGCGGTATGTCGCGG 569  
 DB 577 CAGCGTTCCTATACGCTTGAAGCGCGCGCCCGTATCGCCAAAGG 620

## RESULT 14

AAQ20383  
 ID AAQ20383 standard; DNA; 2214 BP.

AC AAQ20383;  
 XX 27-AUG-2003 (revised)  
 DT 16-APR-1992 (first entry)  
 XX  
 DE ADH complex structural gene (72.000 mol. wt. protein).  
 XX  
 KM Alcohol dehydrogenase; acetic acid; fermentation; sa.  
 XX  
 OS Acetobacter sp.  
 XX  
 PN JF0326988-A.  
 XX  
 PD 27-NOV-1991.  
 XX  
 PF 26-MAR-1990; 90UP-00073440.  
 XX  
 PR 26-FEB-1990; 90UP-00042391.  
 XX  
 PA (NAKA-) NAKANO SUTEN KK.  
 PI Tamaki T, Takemura H, Tayama K, Fukaya M, Okumura H, Kawamura Y;  
 XX  
 DR WPI; 1991-289462/40.  
 DB P-PSDB; AAR20192.  
 XX  
 PT Gene for membrane-bound alcohol dehydrogenase complex - obtd. from  
 PT Acetobacter alcoheligenes, used for prodn. of enzyme for converting  
 PT alcohol to acid.  
 XX  
 PS Disclosure; Fig 3(1-3); 21pp; Japanese.  
 CC The gene encodes a protein of mol. wt. ca. 72.000. Acetobacter  
 CC transformed with the sequence can enhance the efficiency of acetic acid  
 CC fermentation. The ADH complex can be easily extracted from the bacteria  
 CC and purified and it can be used for the determination of an alcohol. See  
 CC also AAQ20383-84, and -86-88. (Updated on 27-AUG-2003 to correct OS  
 CC field.)  
 XX  
 SQ Sequence 2214 BP; 452 A; 683 C; 662 G; 417 T; 0 U; 0 Other;

Query Match 5.0%; Score 87.2; DB 2; Length 2214;  
 Best Local Similarity 50.6%; Pred. No. 1.2e-11;  
 Matches 265; Conservative 0; Mismatches 253; Indels 6; Gaps 2;  
 QY 52 GCGCCCGCGCATTCGCGCAAGGTAAACCGGATTAACGGAAGAACTGTGCGGAACCGGCC 111  
 || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 97 GCATCCGAGATGACGGGCGAGGCGCCACGCGGGGAAGGATCATCCATGCCGATGATCAC 156  
 QY 112 GCTGTGATGATGATTAACATCGGCCGCAACCAAGAAAATATCGCCACTGCCCCCTGACC 171  
 DB 157 CCGCGTAATGATGATACCTATATGCGGCACCTATTTCTGACCAAGCGCTACAGCCCTGGAT 216  
 QY 172 CAGATCATCGCCGCAACAGTTGTCAGTTGCAATGCTGCGGCGCGCGGATGAGAGGCG 231  
 DB 217 CAGATCAACGTTCCAAATGCTGTAACCTGAAGCTGCGCTGCTGATCTGGAACCTGATACC 276  
 QY 232 GGGGCGGTACAGGT---CAGCGCATGATCCATGATGCGGATGATCTGCAAAACCCG 288  
 DB 277 AACCGTGGCCAGGAAGGACGCGCCCTGCTATGATGCGGTGATGATGACGACCAAC 336  
 QY 289 GGTATGATGATCCAGGCGCTGATGCGCAACAGCGCATGATCTGAGGAACACCGCGC 348  
 DB 337 TGGAGCATGATGAAAGCCGTGACGCGCAACCGCAAGCTGTGTGCTATGATGACCG 396  
 QY 349 CAATGCGCCGCGTCCGCAAGCTTAAACGCCCAAGGCGAC---CGCAAGCGCGCGTCCG 405  
 DB 397 CGCGTCCCGCAACATGTCGCGCAAGGCGTGTGATGACGATCAACCGTGGCGGCGCA 456  
 QY 406 CTTTACGGCACAGGCTTATTTCACTCATGGGAACAACATCTGATCGCGCTGATATG 465  
 DB 457 TACTGGAATGGCAAGGCTATTTTCGCAAGTTCGACGGTGCCTGATTCGGCTGACGCC 516  
 QY 466 GAGACGGGCGGAGTTCGATTCGATTCGACGATGCGGCGGAAGACGGCTTGACCAATG 525  
 DB 517 AAGACCGGCAAGCTGTGCTGAGCGTCAACCAATTCCGCCGGAAGCGGAACCTGGGCAAG 576  
 QY 526 AACACCAAGGCGCGGATTTGCGCAATGCGGTATGTCGCGG 569  
 DB 577 CAGCGTTCCTATACGCTTGAAGCGCGCGCCCGTATCGCCAAAGG 620

## RESULT 15

ABA96892  
 ID ABA96892 standard; DNA; 2265 BP.

AC ABA96892;  
 XX  
 XX 07-MAY-2002 (first entry)  
 DT  
 XX  
 DE Gluconobacter oxydans sorbitol dehydrogenase subunit 1 DNA, SEQ ID:6.  
 XX  
 KM Sorbitol dehydrogenase; SDH; subunit 1; Gluconobacter suboxydans;  
 XX  
 OS coenzyme PQQ dependent; haem group; ds.  
 XX  
 GN Gluconobacter oxydans.  
 XX  
 PN KR96069057-A.  
 XX  
 PD 26-OCT-1998.  
 XX  
 PF 26-FEB-1997; 97KR-00005929.  
 XX  
 PR 26-FEB-1997; 97KR-00005929.  
 XX  
 PA (KOAD ) KOREA ADV INST SCI & TECHNOLOGY.  
 XX  
 PI Choi US, Lee SG, Lee EH;  
 XX  
 DR WPI; 1999-608154/52.  
 XX  
 PT Sorbitol dehydrogenase of Gluconobacter suboxydans and gene thereof.  
 XX  
 PS Claim 10; Page 9; 18pp; Korean.

The invention relates to a sorbitol dehydrogenase (SDH) from  
 CC Gluconobacter oxydans (referred to as Gluconobacter suboxydans in the  
 CC specification) and the gene encoding it. The sorbitol dehydrogenase  
 CC comprises 3 subunits of 75 kD, 50 kD and 29 kD, utilises coenzyme PQQ and  
 CC has a haem group. The present sequence represents DNA encoding the 75 kD

CC subunit (subunit 1) of Gluconobacter oxydans sorbitol dehydrogenase  
XX Sequence 2265 BP; 495 A; 670 C; 656 G; 444 T; 0 U; 0 Other;

Query Match 4.9%; Score 85.6; DB 2; Length 2265;  
Best Local Similarity 50.3%; Pred. No. 3.1e-11;  
Matches 304; Conservative 0; Mismatches 279; Indels 21; Gaps 3;

```
QY 100 GCGAACCCCGCGCTGGTGAATGATTAATCACTACGCGCCGCAACCAAGAAAATATATGCCAC 159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 130 GCGAACCCCGCATCCGGTGAATGATGAGCTATAGCCGACCTATTCCAGACGCTAC 189
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 160 TGGCCCTGACCCAGATCACTGCCGACAGCTTGGTCACTGCACTGCTGGCCCGC 219
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 190 AGCCCGCTGGATAGATCACTCAAGGACATGCGAGCACTGAAGTGGCATGGCACTAC 249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 220 GGGATGAGGCGGGGGCCGTACAGTTC--ACCGGATATCCATGATGCGTAT 276
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 250 GATCTGATACCAACCGTGTACAGGAGGTACCCCGCTGATGATGCGCTCATGTAC 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 277 CTGGCAAAACCCCGGTATGATCCAGGCGTGGATCGCAACAGGCGATCTGATCTGG 336
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 310 GCCACCAAACTGAGCAAGATGAAAGCTCTGATGCACTACGGGCAAGCTGCTGG 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 337 GAACACCGCGCCCACTGCGCCCGCTGCGCAAGCTAAACGCCCAAGCGA---CCGCAAG 393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 370 TCTTACGATCCAAAGGTTCCAGGCAACATGCGCACCGGCGTGTGCGATACGGTCAAC 429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 394 CGCGGCGTGCCTTTACGACGACAGCTCTATTTCAGCTCATGGGACAAACCATCTGATC 453
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 430 CGTGTGCAAGCTTACTGAAAGGCAAAAGTCTATTTCGGACCTTCGACGGTGCCTGATT 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 454 GCGCTGATATGAGACGGGCGCAGTCTGATTGATGTC-----GAAAGT 498
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 490 GCCCTGATGCCAAGACCGGCAAGCTGTCTGAGAGCTCTATACGGTCCCAAGGAAGCG 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 499 GGATCGGCGCAAGACGGCTTGACCACTAACACACCGGGCGCATGTGCGCAATGGCGTC 558
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 550 CAGCTGGTCAACGAGCTCTCTACACGTTGACGCTGCCGTTATGCCCAAGGGCAAG 609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 559 ATGCTGCGGGGTTCCACTGCCAATATTGCCCTATGATGCTTTATCTCGGGGCAAGAT 618
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 610 GTATCTCTCGGCAACGGCGGTGCAAGTTCGGCGCCCGCTGGCTTCGTGACGGGTATGAC 669
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 619 TCCGCGACGGGTAGAGAGCTGTGGCGCAACACTTTATCCGCAAGCGGCGAAGGCT 678
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 670 GCTGAAACGGGAAAGATGACTGGCGCTTCTTACCCGTTCCGAACCTTGACAAACAGCCG 729
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 679 GACG 682
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 730 GACG 733
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: November 22, 2004, 14:45:01  
Job time : 949 secs



```

Db      301 CAGGCGCTGATGCGCAACAGGCGATCTGATCTGGGAAACACGCGCCCACTGCCCCC 360
Qy      361 GTGCGACGCTAAACGCCCAAGGCGACCGCAAGCGCGCGTGCCTTTTACGCGACGAGC 420
Db      361 GTGCGACGCTAAACGCCCAAGGCGACCGCAAGCGCGCGTGCCTTTTACGCGACGAGC 420
Qy      421 CTCTATTTTGACTCATGCGCAACCATCTGATGCGCTGGAATGGAACGCGGCCAGGTC 480
Db      421 CTCTATTTTGACTCATGCGCAACCATCTGATGCGCTGGAATGGAACGCGGCCAGGTC 480
Qy      481 GATTCGATGTCGAGTGGATGCGGCGAAGACGCGCTTGACCACTAACACGCGGCGCG 540
Db      481 GATTCGATGTCGAGTGGATGCGGCGAAGACGCGCTTGACCACTAACACGCGGCGCG 540
Qy      541 ATTGTCGCAATGCGCTCATCTGCGCGGTTCCACTGCAATATTGCGCTTATGGAATGC 600
Db      541 ATTGTCGCAATGCGCTCATCTGCGCGGTTCCACTGCAATATTGCGCTTATGGAATGC 600
Qy      601 TTTATCTCGGGGACGATTCGCGGAGGAGGAGCTGCGGCGCAACCACTTTATCCG 660
Db      601 TTTATCTCGGGGACGATTCGCGGAGGAGGAGCTGCGGCGCAACCACTTTATCCG 660
Qy      661 CAGCGCGGCGAAGAGGAGTGAAGACCTTGGGCAATGTTTGAAGCGCGCTGATGAC 720
Db      661 CAGCGCGGCGAAGAGGAGTGAAGACCTTGGGCAATGTTTGAAGCGCGCTGATGAC 720
Qy      721 GGGGCTCTGGGCTGATCATCTATGATCCCTGACCACTTGTGTTCTATGCTGACG 780
Db      721 GGGGCTCTGGGCTGATCATCTATGATCCCTGACCACTTGTGTTCTATGCTGACG 780
Qy      781 GGGGCTGGGCGCAGCTGCGCAACCCAGCGCGGCGCGCGCGGCGGCGCTGATGAC 840
Db      781 GGGGCTGGGCGCAGCTGCGCAACCCAGCGCGGCGCGCGCGGCGGCGCTGATGAC 840
Qy      841 AACACCGGCTTTCGCGTGCCTCCGACAGCGGCGGAGATTTGCTGCGGCTACCAAGCCTG 900
Db      841 AACACCGGCTTTCGCGTGCCTCCGACAGCGGCGGAGATTTGCTGCGGCTACCAAGCCTG 900
Qy      901 CCGCGGAGCACTGGGAGCAAGATGCACTTGAATGATGATGATGATGATGATGATG 960
Db      901 CCGCGGAGCACTGGGAGCAAGATGCACTTGAATGATGATGATGATGATGATGATG 960
Qy      961 CAAACCTCGGCGGAGTGAAGGCTGCGCGCATCAACCCCAATGCGCGGCGAG 1020
Db      961 CAAACCTCGGCGGAGTGAAGGCTGCGCGCATCAACCCCAATGCGCGGCGAG 1020
Qy      1021 CGCGGCTGCTGAAGGCTGCGCTTGCAGACCGGAGCGATGCGTGTGATGCGGCG 1080
Db      1021 CGCGGCTGCTGAAGGCTGCGCTTGCAGACCGGAGCGATGCGTGTGATGCGGCG 1080
Qy      1081 TCGGCGCAATTCCTGCTGGGCGCGTGAATCCCACTACCAATATGATGCTGATGAC 1140
Db      1081 TCGGCGCAATTCCTGCTGGGCGCGTGAATCCCACTACCAATATGATGCTGATGAC 1140
Qy      1141 GAGACCGGCTTGTGAAGGAGGAGATGCGGCTGGAAGAGTGAAGCTGATGAT 1200
Db      1141 GAGACCGGCTTGTGAAGGAGGAGATGCGGCTGGAAGAGTGAAGCTGATGAT 1200
Qy      1201 GACGCTGCGCGACCTTCTGGGAGGCGGAGCTGCTGCAAGCCGCACTGAAACCGGAC 1260
Db      1201 GACGCTGCGCGACCTTCTGGGAGGCGGAGCTGCTGCAAGCCGCACTGAAACCGGAC 1260
Qy      1261 ACCGCGCATTTACTTCTGCGCTGGAACAATGCTGCTGCAATATATGATGATGAT 1320
Db      1261 ACCGCGCATTTACTTCTGCGCTGGAACAATGCTGCTGCAATATATGATGATGAT 1320
Qy      1321 GAGTTTACGCGCTGCAAGCTTAAACACAGGCGGCAACCGCAAACTGCGCGCGCTTT 1380
Db      1321 GAGTTTACGCGCTGCAAGCTTAAACACAGGCGGCAACCGCAAACTGCGCGCGCTTT 1380
Qy      1381 GAAATATGCGCGCATGCAAGCGGATGATATGACACCGGCGCACTTGTGCTGCGG 1440
Db      1381 GAAATATGCGCGCATGCAAGCGGATGATATGACACCGGCGCACTTGTGCTGCGG 1440

```

```

Qy      1441 GAGCGCTTGGCGGAACTATCTGCGCCCTTTTGTGCAAGCGGCGGCTGTGTTCAAC 1500
Db      1441 GAGCGCTTGGCGGAACTATCTGCGCCCTTTTGTGCAAGCGGCGGCTGTGTTCAAC 1500
Qy      1501 GCGCGGACCGACGCTATTCTCGTGCCTGACGCAAGAAACCGCGCACTTTTGTGCGAG 1560
Db      1501 GCGCGGACCGACGCTATTCTCGTGCCTGACGCAAGAAACCGCGCACTTTTGTGCGAG 1560
Qy      1561 GCCCGTCTTGCAGCGGTCGCGAGCGGCGGCGGATGAGCTGACAGTTGACGCGCTGCA 1620
Db      1561 GCCCGTCTTGCAGCGGTCGCGAGCGGCGGCGGATGAGCTGACAGTTGACGCGCTGCA 1620
Qy      1621 TATATGCGCATGCGGTCGCGGCGGCTGACCTTATGCAACGCAATTTGAACGCGCGTGC 1680
Db      1621 TATATGCGCATGCGGTCGCGGCGGCTGACCTTATGCAACGCAATTTGAACGCGCGTGC 1680
Qy      1681 GAGGCAATGATTCGACCTGCGTGTGATGCAATGATGATGATGATGATGATGATG 1740
Db      1681 GAGGCAATGATTCGACCTGCGTGTGATGCAATGATGATGATGATGATGATGATG 1740

```

```

RESULT 2
US-09-470-667-1
; Sequence 1, Application US/09470667
; Patent No. 6730503
; GENERAL INFORMATION:
; APPLICANT: Asakura, Akira
; APPLICANT: Hoshino, Tatsuo
; APPLICANT: Ojima, Setsuko
; APPLICANT: Shinjoh, Masako
; APPLICANT: Tomiyama, No. 6730503ibumi
; TITLE OF INVENTION: No. 6730503el Alcohol/Aldehyde Dehydrogenases
; FILE REFERENCE: C38435/109700CON
; CURRENT APPLICATION NUMBER: US/09/470,667
; PRIOR APPLICATION NUMBER: US 08/934,506
; PRIORITY FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1740
; TYPE: DNA
; ORGANISM: Gluconobacter oxydans
US-09-470-667-1

```

```

Query Match      64.9%; Score 1128.6; DB 4; Length 1740;
Best Local Similarity 78.8%; Pred. No. 5.3e-288;
Matches 1373; Conservative 0; Mismatches 364; Indels 6; Gaps 2;

```

```

Qy      1 ATGAACCCCAACAGCTGCTTTCGACCAAGCGCGCGTGTCTATTGCTTACCGGCGCGCC 60
Db      1 ATGAACCCCAACAGCTTGTGCTTTCGACCAAGCTGCGCACTTGTGCTTGTGCGCACCC 60
Qy      61 GCATTGCGGAGGTAACCCGATTAACGATGAACCTGCGGCAACCCGCGCTGTGTA 120
Db      61 GCATTGCTCAAGTGAACCCCGTCAACCGATGAATGCTGCGCAACCCGCGCTGTGTA 120
Qy      121 TGAATTACTAGCGCGGCAACGAAGAACTATATGCGCATCGCCCTGACACCGATGAT 180
Db      121 TGAATAGCTAGCGTGAACCAAGAACTATGCTGCGGCGCGCGGCAAGCAAGTATG 180
Qy      181 GCCGACAACTGCTGCTGATGCAACTGCTGCGGCGCGGAGTGAAGGCGGCGGCGTA 240
Db      181 ACTGAGAACTGCGGCAACCTGCAACTGCTGCGGCGCGGCGGCAAGTATG 240
Qy      241 CAGTCAACCGCGATGATCATGATGCGGATGATGATGATGATGATGATGATGATGATG 300
Db      241 CAGTCAACCGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
Qy      301 CAGGCGCTGATGCGGCAACAGGCGATGATGATGATGATGATGATGATGATGATGATG 360
Db      301 CAGGCGCTGATGCGGCAACAGGCGATGATGATGATGATGATGATGATGATGATGATG 360

```

QY 361 GTGCCACGCTAAACGCGCAAGGCGACCGGCAAGCGGCGTCCCTTTACGCGACAGC 420  
 Db 361 ATGGCCACGCTGAACAGCTTTGGCGAGCCGACCCCGGCGATGGCGCTGTACCGACACAC 420  
 QY 421 CTCTATTTCAGCTCATAGGACCAACATCTGATCGCGCTGATATGGAACGCGCCAGTTC 480  
 Db 421 GTTACTCTTTGTTCGCGGACCAACACCTGCTGCGCTCGACACCCGCACTGCGCCAGTTC 480  
 QY 481 GTATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540  
 Db 481 AGCTTGAAGCTGACCGCGCCGCAAGGCGAAGACATGCTTTCG--AACTGTCGCGCCG 537  
 QY 541 ATTGTCGCAATGCGCTCATGCTGCGGCTTCGACCTGCGCAATATTCGCTTATGATGAC 600  
 Db 538 ATGTCGCAAAACGCGCTGATGCTGTCGCGCTTCGACCTGCGCAATATTCGCGCTGCTG 597  
 QY 601 TTTATCTCGGGGAGACATTCGCGGACGCGGTGAGAGAGCTGTGGGCGAACCACTTTATCCG 660  
 Db 598 TTTGCTCGGGCCACACTCGCGCACCGGTGAGAGAGCTGTGGGCGAACCACTTTATCCG 657  
 QY 661 CAGCCGCGGAGAGAGGATGACGAGACTTGGGCGAATGATTTGAGGCGCGCTGATGAC 720  
 Db 658 CGGCTGGCGAAGAGGATGATGAGACTTGGGCGAAGATTTACGAGCCGCTTGAATGAC 717  
 QY 721 GCGCTCTGGGGTCAATCACTATGATCCCGTACGACACCTTGTCTTATGAGCTTCGAC 780  
 Db 718 GGTGCTCGGGCGCAATCACTATGACCCCGTACACACCTTGTCTTATGAGCTTCGAC 777  
 QY 781 GCGCTGGGCGCGAGTCCGAAACCGACGCGCGGCGCGCGCGCGCGCGCTGATGAGAC 840  
 Db 778 GCTGTGGGCTCGCGCTCGGAAACCGACGCGCGCGCGCGCGCGCGCTGATGAGAC 837  
 QY 841 AACACCGCTTTGCGGTGCTGCGGACGCGGAGATGCTGCGGTACCGACGACCTG 900  
 Db 838 AACACCGCTTTGCGGTGCTGCGGACGCGGAGATGCTGCGGTACCGACGACCTG 897  
 QY 901 CCGCGGAGACACTGGAACCAAGATGACGCTTGAATGATGATGATGATGATGATGATGATG 960  
 Db 898 CCGCGGAGACACTGGAACCAAGATGACGCTTGAATGATGATGATGATGATGATGATG 957  
 QY 961 CAACTCTGCGCGAGATGAGAGGCTGCGCGCATCAACCCCAATGCGCGGACGCGCGAG 1020  
 Db 958 CAACTCTGCGCGAGATGAGAGGCTGCGCGCATCAACCCCAATGCGCGGACGCGCGAG 1017  
 QY 1021 CCGCGGTGCTGACGCGGTGCGCTTGAAGACCGGACGATGCTGCTTGAATGAGCGC 1080  
 Db 1018 CCGCGGTGCTGACGCGGTGCGCTTGAAGACCGGACGATGCTGCTTGAATGAGCGC 1077  
 QY 1081 TCGGCGGAATTCCTGTCGCGCGGCTGATACCACTACCAATATGATGCTGATGAC 1140  
 Db 1078 ACCGCGGAATTCCTGTCGCGCGGCTGATACCACTACCAATATGATGATGATGAC 1137  
 QY 1141 GAGACCGCGCTTGTGACGCTGAACGAGATGCGGTCTGAAGAAGCTGACGCTGAATAT 1200  
 Db 1138 GAGACCGCGCTTGTGACGCTGAACGAGATGCGGTCTGAAGAAGCTGATGATGATAT 1197  
 QY 1201 GAGCTGCGCGGACCTTCTGCGGTGCGCGGCGGATGCTGCTGACGCGGACCTGAACCGGAC 1260  
 Db 1198 GAGCTGCGCGGACCTTCTGCGGTGCGCGGCGGATGCTGCTGACGCGGACCTGAACCGGAC 1257  
 QY 1261 ACCGCGATTTACTTCTTGGCGCTGAACGATGCTGCTGATGATGATGATGATGATGATG 1320  
 Db 1258 AGCGGATCTACTTCTTGGCGCTGAACGATGCTGCTGATGATGATGATGATGATGATG 1317  
 QY 1321 GATTTAGCGCGCTGACGCTTATACACGAGCGACCGCAAACTGCGCGCGGCTTT 1380  
 Db 1318 GATTTAGCGCGCTGACGCTTATACACGAGCAACGATGCTGCGCGCGGCTTT 1377  
 QY 1381 GAAATATGCGCGCGATGACGCGGATGATGATGATGATGATGATGATGATGATGATGATG 1440  
 Db 1378 GATATGATGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1437

QY 1441 GAGCGCCCTGCGCGGAATCTACCTCCGCTTTGTGACGCGGAGCGGCTGTGATCAAC 1500  
 Db 1438 GAGCTGCTGCGCGGAATCTATTCGCGCTTGTGACGCGGAGCGGCTGTGATCAAC 1497  
 QY 1501 GCGCGGACGACCGCTATTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCT 1560  
 Db 1498 GGTGCTGAGATCTCTTCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCT 1557  
 QY 1561 GCGCGCTTCCGCGAGGTGCGGACGCGGAGCGGATGACGCTGATGATGATGATGATGATGAT 1620  
 Db 1558 ACCGCGCTTCCGCGAGGTGCGGAGCGGATGACGCTGATGATGATGATGATGATGATGAT 1617  
 QY 1621 TATATGCGCATGCGTGGCGGCGGCTGATGATGATGATGATGATGATGATGATGATGAT 1679  
 Db 1618 TATGTCGCGATGCGAGGCGGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1677  
 QY 1680 --CGAGGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 1737  
 Db 1678 GCGAGCGAGTGCATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 1737  
 QY 1738 TAA 1740  
 Db 1738 TAA 1740

RESULT 3  
 US-09-470-667-3  
 ; Sequence 3, Application US/09470667  
 ; Patent No. 6730503  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Asakura, Akira  
 ; APPLICANT: Hoshino, Tatsuo  
 ; APPLICANT: Ojima, Setsuko  
 ; APPLICANT: Shinjoh, Masako  
 ; APPLICANT: Tomiyama, No. 6730503;Dum1  
 ; TITLE OF INVENTION: No. 6730503el Alcohol/Aldehyde Dehydrogenases  
 ; FILE REFERENCE: C38435/109700CON  
 ; CURRENT APPLICATION NUMBER: US/09/470,667  
 ; CURRENT FILING DATE: 1999-12-22  
 ; PRIOR APPLICATION NUMBER: US 08/934,506  
 ; NUMBER OF SEQ ID NOS: 12  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 3  
 ; LENGTH: 1737  
 ; TYPE: DNA  
 ; ORGANISM: Gluconobacter oxydans  
 ; US-09-470-667-3

Query Match 62.0%; Score 1078.2; DB 4; Length 1737;  
 Best Local Similarity 77.1%; Pred. No. 1e-274;  
 Matches 1340; Conservative 0; Mismatches 393; Indels 6; Gaps 2;

QY 1 ATGAACCCCAACAGCTGCTTGCACACGCGCGCGCTGATGCTTACCGCGCGCC 60  
 Db 1 ATGAACCTGACGACCTGCTGCAAGACGCGCGCGCTGCTGCTTGCACACATTC 60  
 QY 61 GCATTTCGCGGAGTAACCCCGATTAACGATGAACTGCTGCGGAAACCGCGCGCTGCGGAA 120  
 Db 61 GCGCTTCCCA--AACCGCATCAACGATGAATGCTGCGGAAACCGCGCGCTGCGGAA 117  
 QY 121 TGAATTAATGACGCGCAACCAAGAAATATGCACTGCGCTGACCGCATCACT 180  
 Db 118 TGAATTAATGACGCGCAACCAAGAAATATGCACTGCGCTGACCGCATCACT 177  
 QY 181 GCGCAACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240  
 Db 178 GCAACAAACGTCGCGCAACGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 237  
 QY 241 CAGGTCAACGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
 Db 238 CAATGACCGCGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 297

QY 301 CAGCGCTGATGATGCGCAAAAGGCGATCTGATCTGGGAAACACCGCGCACTGCGCCG 360  
 Db 298 CAGGCGATGAGCGCGCGGCGATCTGATCTGGGAAACACCGCGCACTGCGCCG 357  
 QY 361 GTGGCCACGTAAAGCGCCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420  
 Db 358 ATGCGACGCTGAACGCTTTGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 417  
 QY 421 CTCTATTTGAGCTCATGAGCAACACTGATGCGCGCTGATATGAGAGCGGCGGCGG 480  
 Db 418 GTCTATTTGGTCTCGTGGGCAACACTGATGCGCGCTGAGACCGCGGCGGCGGCGG 477  
 QY 481 GTATTGAGTGCAGAGGTGATGGCGGCAAGAGGCTTACAGATTAACACCGGCGGCG 540  
 Db 478 GTATTGAGTGCAGAGGTGATGGCGGCAAGAGGCTTACAGATTAACACCGGCGGCG 534  
 QY 541 ATGTGCGCAATGGCGCTCATGCGCGGCTGCGCGGCTGCGCAATTTGCGCGCTGATG 600  
 Db 535 ATGTGCGCAATGGCGCTCATGCGCGGCTGCGCGGCTGCGCAATTTGCGCGCTGATG 594  
 QY 601 TTTATCTCGGCGGCGAGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660  
 Db 595 TTGCTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 654  
 QY 661 CAGCGCGGCGGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720  
 Db 655 CCGCGCGGCGGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 714  
 QY 721 GCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780  
 Db 715 GCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 774  
 QY 781 GCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840  
 Db 775 GCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 834  
 QY 841 AACACCGGCTTTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900  
 Db 835 AACACCGGCTTTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 894  
 QY 901 CCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960  
 Db 895 CCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 954  
 QY 961 CAGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020  
 Db 955 CAGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1014  
 QY 1021 CCGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080  
 Db 1015 CCGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1074  
 QY 1081 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1140  
 Db 1075 ACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1134  
 QY 1141 GAGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1200  
 Db 1135 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1194  
 QY 1201 GAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1260  
 Db 1195 GAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1254  
 QY 1261 ACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1320  
 Db 1255 ACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1314  
 QY 1321 GAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380  
 Db 1315 GAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1374  
 QY 1381 GAAATATAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440

Db 1375 GAAATATAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1434  
 QY 1441 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1500  
 Db 1435 GAAATATAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1494  
 QY 1501 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1560  
 Db 1495 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1554  
 QY 1561 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1620  
 Db 1555 ACCGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1614  
 QY 1621 TATATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1680  
 Db 1615 TATATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1674  
 QY 1681 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1739  
 Db 1675 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1733

RESULT 4  
 US-09-470-667-2  
 ; Sequence 2, Application us/09470667  
 ; Parent No. 6730503  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Asakura, Akira  
 ; APPLICANT: Hoshino, Tatsuo  
 ; APPLICANT: Ojima, Setsuko  
 ; APPLICANT: Shinjoh, Masako  
 ; APPLICANT: Tomiyama, No. 6730503ibumi  
 ; TITLE OR INVENTION: No. 6730503el Alcohol/Aldehyde Dehydrogenases  
 ; FILE REFERENCE: C38435/109700CON  
 ; CURRENT FILING DATE: US/09/470,667  
 ; PRIOR FILING DATE: 1999-12-22  
 ; PRIOR FILING DATE: 1997-09-19  
 ; NUMBER OF SEQ ID NOS: 12  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 2  
 ; LENGTH: 1740  
 ; TYPE: DNA  
 ; ORGANISM: Gluconobacter oxydans  
 ; US-09-470-667-2

Query Match 60.6%; Score 1053.6; DB 4; Length 1740;  
 Best Local Similarity 75.7%; Pred. No. 3.2e-268;  
 Matches 1318; Conservative 0; Mismatches 419; Indels 3; Gaps 1;

QY 1 ATGAACCCCAACGCTGCTTGGCAACGCGCGCGTGTCTATTGCTTACCGGCGCGC 60  
 Db 1 ATGAAGAGCTGCTTGTGCTGCTTGGGAGCGTGGCGGCTTGAAGCTTACCTCTT 60  
 QY 61 GCATTGCGGAGGTAACCCGATTTACGATGAAGTGTGCGGAACCGCGCGGCGGTA 120  
 Db 61 GCGGTTCTCAAGTACCCCGTCCACCGATGAATGCTGCGGAACCGCGCGGCGGTA 120  
 QY 121 TGAATTAAGTACGCGCGCAACCAAGAAATCTATGCGCATGCGCGCTGACCCAGATCAT 180  
 Db 121 TGAATTAAGTACGCGCGCAACCAAGAAATCTATGCGCATGCGCGCTGACCCAGATCAT 180  
 QY 181 GCCGACAAGTGGTCACTGTTGCACTGCTGCGCGCGCGGAGTGAAGGCGGCGCGTA 240  
 Db 181 ACTGAGAACGTCGCGCAACGCAACCTGCTGCGCGCGCGGAGTGAAGGCGGCGCGTA 240  
 QY 241 CAGGCTACGCGGATGATCAATGATGCGGCTGATGATCTGCAACACCGGCTGATGATC 300  
 Db 241 CAGGCTACGCGGATGATCAATGATGCGGCTGATGATCTGCAACACCGGCTGATGATC 300  
 QY 301 CAGGCGCTGATGCGCAACAGCGGATGATCTGATCTGGGAACACCGCGCGCACTGCCGCC 360

Db 301 CAGGCCATCAGACCAAACTGGCGATCTGATCTGGGACACCGCGCCCACTGCGCAAC  
 Qy 361 GTGGCCACGCTTAAACCGCCCAAGCGACCGCAAGCGCGCGTGGCCCTTAAACGACGACG  
 Db 361 ATGCGCACGCTGAACAGCTTTGGGAGACCGACCGCGCGCTGTAAGGACACAC 420  
 Qy 421 CTCTATTTCAGCTCATAGGACAAACCATCTGATCGCGTGGATATGAGACGCGCCAGCTC  
 Db 421 GTTACTTTTGTTCGTGGAGCAACCACTGGTCCCTCGACACCGCAATGSCCAAGTG 480  
 Qy 481 GTATTGATTCGAAAGCTGATCGGGGAGAGACGCGTTCAGCACTAACACAGCGGGCGG  
 Db 481 ACGTTCAGCTGACCGCGCCCAAGGAGAGACATGGTTTCG--AACTGTCGCGCCG 537  
 Qy 541 ATTGCGCCAAATGAGCGTATCGTCCGCGGTTCCACCTGCCAATATTCGCGCTATGATG  
 Db 538 ATGCGGCMAACCGCGGATCGTTGCCGTTTCGACCTGCCAATACCTGCGGCTGCG 597  
 Qy 601 TTTATCTCGGGGACAGATTCGCGGACGGGTGAGAGAGCTGTGCGCAACCATTTATCCG  
 Db 598 TTTGCTCGGGGACAGATTCGCGGACCGGTGAGAGAGCTGTGCGCAACCATTTATCCG 657  
 Qy 661 CAGCGCGGACGAGAGGGGTGACGAGCTTGGGGCAATGATTCGAGGGCGCGTGGATGAC 720  
 Db 658 CGCGCTGCGAAGAGGTGATGAGACTTGGGGCAACGATTCAGAACCGCGTTGATGAC 717  
 Qy 721 GCGCTCTGGGGTCAATCATATGATCCGCTGACGAACTTGTGTTCTATGCGTCAAC 780  
 Db 718 GCGCTCTGGGGTCAATCATATGATCCGCTTGGCGGCTTGTCACTGCGGCTGCTG 777  
 Qy 781 GCGCTGGGGCCCAAGCTCCGAAACCGAGCGGCGGACCGCGCGGCGCTGATGCGAC 840  
 Db 778 GCTGTTGCGCGGCTTGGAAACCGAGCGGCGACACCGCGCGACCATGACGAC 837  
 Qy 841 AACACCGCTTGGCGGTGCGTCCGACACCGGCGGAGATTGTCGCGTCAACGACCTG 900  
 Db 838 AACACCGCTTGGCGGTGCGTCCGACACCGGCGGAGTCTGCGGCTCAACCACTG 897  
 Qy 901 CCGCGGACAACTGGGACCAAGATGACGTTGAGATGATGCTCCGCAACGTCGATG 960  
 Db 898 CCGCGGACAACTGGGACCAAGATGACGTTGAGATGATGCTCCGCAACGTCGATG 957  
 Qy 961 CAACCTCGGCGGATGAGAGGCTCGCGGACCAACCCGATGCGGCGGACGCGGAC 1020  
 Db 958 CAGCGCGGACGATGACGAGCGGCTCCGCTGATCAACCGGACCGCGGACGCGGAC 1017  
 Qy 1021 CGCGGTGCTGACGCGGTGCGCTTGCAGACCGGACGATGCTGCTTGTGATGCGG 1080  
 Db 1018 CGTGCCTGCTGACCGCGGTGCGCTTGCAGAACCGGACGATGCGGACGTCGAC 1077  
 Qy 1081 TCGGGGCAATTCCTGCTGGGCGCGGATACCACTACCACTAATGATGCGCTGATGAC 1140  
 Db 1078 ACCGCGCAATTCCTGCTGGGCGCGGATACCACTACCACTAATGATGATGAC 1137  
 Qy 1141 GAGACCGGCTTGTGACGCTGAAACGAGATCGGTGCTGAAAGAGCTGACGTTGAAT 1200  
 Db 1138 GAAACCGGATCGTACGCTGACGAGTCAAGTTCTGACGAGCTGACACCCCTAT 1197  
 Qy 1201 GACGCTCGGCGGACCTTCTGCGGCGGCGGACGCTGCTGACCGGACGACCGGAC 1260  
 Db 1198 GACGCTCGGCGGACCTTCTGCGGCGGCGGACGCTGCTGACCGGACCGGAC 1257  
 Qy 1261 ACCGCGATTTACTCTTCTGCGGCGGCAATGCTGCTACATATTTATGCGCGTGTCA 1320  
 Db 1258 ACCGCGATTTACTCTTCTGCGGCGGCAATGCTGCTACATATTTATGCGCGTGTCA 1317  
 Qy 1321 GAGTTTACCGGCTGACGCTGATTAACAACGCGGACCGGCAAACTCGCGCGGCTTT 1380  
 Db 1318 GAAATTCGCTGCGGAGCTGTAACAACGAGCTGACCGGCAAAAGGACCGGCTTAA 1377  
 Qy 1381 GAAATTCGCTGCGGAGCTGTAACAACGAGCTGTAACAACGAGCTGACCGGCTTAA 1440

Db 1378 GAGCTGTGGCGGCTATGACGCCATGACATCAGACAGGCGCGACCTGTGACCGCT 1437  
 Qy 1441 GAGCGCTTGGCGGCAACTACTGCGCCGTTTGTGACAGCGGCGGCTGTGTTCAAC 1500  
 Db 1438 GAGCGGCAAGCTCGAATTAAGCGGCTGTCTGTGACCGGCTGGCGGCTTCTTCAAC 1497  
 Qy 1501 GCGGGAACCGGCTATTTTCCGTGCGCTTCCAGCGGAAACCGGCGGACCTTTGTGAC 1560  
 Db 1498 GCGGGAACCGGCTATTTTCCGTGCGCTTCCAGCGGAAACCGGCGGACCTTTGTGAC 1557  
 Qy 1561 GCGGCTTGGGACGCGTGGCGGACGCGGCGGAGATGACGATGAGATTGAGCGGCGAA 1620  
 Db 1558 ACCGCTTGGGACGCGTGGCGGACGCGGCGGAGATGACGATGAGATTGAGCGGCGAA 1617  
 Qy 1621 TATATGCGCATCGTGGCGGCGGCTGACCTATGCGACGCAATTTAAACCGCGCGTGGCC 1680  
 Db 1618 TACATGCGCATCGGCGGCGGCGGCGGACGCACTATGCTTCCACAAACCGCTCGCGC 1677  
 Qy 1681 GAGGCAATGCAATGCACTCGTGGCTGATATGCGATGCTTGTGACGCGGCGGATTA 1740  
 Db 1678 GAGCGGTGACTGACCGCGGATGCGATGCTTGTGCGGCTGCCCGACGA 1737

RESULT 5  
 US-07-985-458-1  
 ; Sequence 1, Application US/07985458  
 ; Patent No. 5344777  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tamaki, Toshimi;  
 ; APPLICANT: Takemura, Hiroshi;  
 ; APPLICANT: Takemura, Kenji;  
 ; APPLICANT: Fukaya, Masahiro;  
 ; APPLICANT: Okumura, Hajime and  
 ; APPLICANT: Kawamura, Yoshiya  
 ; TITLE OF INVENTION: Structural Gene of Membrane-Bound  
 ; TITLE OF INVENTION: Alcohol Dehydrogenase Complex, Plasmid  
 ; TITLE OF INVENTION: Containing The Same And Transformed Acetic Acid  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Frisshauf, Holtz, Goodman & Woodward, P.C.  
 ; STREET: 600 Third Avenue  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10016-2088  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5 inch, 0.72 mb  
 ; COMPUTER: IBM PC compatible (NEC PC-9801 ES)  
 ; OPERATING SYSTEM: MS DOS  
 ; SOFTWARE: ASCII Form  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/985,458  
 ; FILING DATE: 19921203  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/658,221  
 ; FILING DATE: 20-FEB-1991  
 ; APPLICATION NUMBER: 73440/1990  
 ; FILING DATE: 26-MAR-1990  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Goodman, Herbert  
 ; REGISTRATION NUMBER: 17081  
 ; REFERENCE/DOCKET NUMBER: 910134/HG  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212)972-1400  
 ; TELEFAX: (212)370-1622  
 ; TELEX: 236268  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2214 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double

```

; TOPOLOGY: unknown
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Acetobacter alioacetigenes
; STRAIN: MH-24
; PUBLICATION INFORMATION:
; AUTHORS: Tamaki, Toshimi;
; AUTHORS: Fukaya, Masahiro;
; AUTHORS: Takemura, Hiroshi;
; AUTHORS: Tayama, Kenji;
; AUTHORS: Okumura, Hajime;
; AUTHORS: Kawamura, Yoshiva;
; AUTHORS: Nishiyama, Makoto;
; AUTHORS: Horinouchi, Suenaru and
; AUTHORS: Beppu, Teruhiko
; TITLE: Cloning and Sequencing of the Gene Cluster
; TITLE: Encoding Two Subunits of Membrane-Bound
; TITLE: Alcohol Dehydrogenase from Acetobacter
; JOURNAL: Biochimica et Biophysica Acta.
; VOLUME: 1088
; PAGES: 292-300
; DATE: 1991
; US-07-985-458-1

```

```

Query Match      5.1%; Score 88.8; DB 1; Length 2214;
Best Local Similarity 50.8%; Pred. No. 1e-13;
Matches 266; Conservative 0; Mismatches 252; Indels 6; Gaps 2;

```

```

QY 52 GCGCCCGCCGATTCGCGAGTAAACCCGATTACCGATGAACTGTGCGAAACCCGCC 111
DB 97 GCATCCGAGATGACGGGCGAGCGCCACGCGGGAAAGCATATCATGCCGATGATC 156
QY 112 GCTGCGAATGATTAATCTACGCGCGGCAACAGAAATATGACCATGCGCCCGTGA 171
DB 157 CCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 216
QY 172 CAGATCACTGCGCAACAGTGTGATGATGATGATGATGATGATGATGATGATG 231
DB 217 CAGATCAACGCTTCCATATGATGATGATGATGATGATGATGATGATGATGATG 276
QY 232 GGGGCCGTACAGT--CAGCGCGATGATCATGATGATGATGATGATGATGATGAT 288
DB 277 AACCGTGCGCAGAGAGGACAGCGCCCTGTATTGATGATGATGATGATGATGAT 336
QY 289 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 348
DB 337 TGAAGCATGATGAAGCCGTCGACGCGCAACCGCAAGCTGTGTGTCTATATGAC 396
QY 349 CAATGCGCGCGCTGCGCACTAAACGCCCAAGCGAC--CGCAAGCGCGCGTCC 405
DB 397 CCGGTCGCGCGCAACATGCGCAAGGCGCTGTGTGTGTGTGTGTGTGTGTGTGT 456
QY 406 CTTCAGCGCAAGCGCTTATTCAGTCAATGGAACACCATGATGATGATGATGAT 465
DB 457 TACTGGAATGCAAGGCTTATTCAGTCAATGGAACACCATGATGATGATGATGAT 516
QY 466 GAGAAGCGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 525
DB 517 AAGACCGCGCAAGCTGCTGGAAGCTCAACCATTCGCGCGCGCAAGCGAACTG 576
QY 526 AAGACCGCGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATG 569
DB 577 CAGCGTTCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 620

```

```

RESULT 6
US-09-296-284-22
; Sequence 22, Application US/09296284A
; Patent No. 6204040
; GENERAL INFORMATION:
; APPLICANT: Choi, Eun-Sung
; APPLICANT: Rhee, Sang-Ki

```

```

; APPLICANT: Lee, Eun-Hae
; TITLE OF INVENTION: Gluconobacter Suboxydans Sorbitol Dehydrogenase, Genes
; TITLE OF INVENTION: and Methods of Use Thereof
; FILE REFERENCE: 153,087000
; CURRENT APPLICATION NUMBER: US/09/296,284A
; CURRENT FILING DATE: 1999-04-22
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 22
; LENGTH: 2163
; TYPE: DNA
; ORGANISM: Gluconobacter suboxydans
; US-09-296-284-22

```

```

Query Match      4.9%; Score 85.6; DB 3; Length 2163;
Best Local Similarity 50.3%; Pred. No. 7e-13;
Matches 304; Conservative 0; Mismatches 279; Indels 21; Gaps 3;

```

```

QY 100 GGAACCCGCGCGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 159
DB 28 GCGACCGCATCCGCGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 87
QY 160 TCGCCCTGACCCAGATCACTGCGCAACGTTGTGATGATGATGATGATGATGAT 219
DB 88 AGCCGCTGATCAGATCAACCAAGACATGCGAATCTGAAGCTGGCATGGCACTAC 147
QY 220 GGAATGAGAGGCGCGCGTACAGTCC--ACGCGATGATTCATGATGATGATGATGAT 276
DB 148 GATCTGATCAACCGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 207
QY 277 CTGGCAACCCCGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 336
DB 208 GCGACCAACATGAGCAAGATGAAAGCTGTGATGATGATGATGATGATGATGATG 267
QY 337 GAAACCGCGCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 393
DB 268 TCTTACATCCAAAGTTTCCAGCAACATGCGCAACCGCGCGCGCGCGCGCGCG 327
QY 394 GCGCGCGCGCGCTTACAGCGCAAGCGCTTATTCAGTCACTGATGATGATGATGAT 453
DB 328 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 387
QY 454 GCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 498
DB 368 GCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 447
QY 499 GATCGCGCGCAAGCGCTTACAGCAACATGCGCAACCGCGCGCGCGCGCGCGCG 558
DB 448 CAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 507
QY 559 ATGCTGCGCGGTTCACCTGCAATGATGATGATGATGATGATGATGATGATGAT 618
DB 508 GTTCATATCCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 567
QY 619 TCGCGACCGGTTGAGAGCTGTGCGCAACCATTTATCCGCGCGCGCGCGCGCG 678
DB 568 GCTGAACGGGAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 627
QY 679 GAGC 682
DB 628 GAGC 631

```

```

RESULT 7
US-09-296-284-1
; Sequence 1, Application US/09296284A
; Patent No. 6204040
; GENERAL INFORMATION:
; APPLICANT: Choi, Eun-Sung
; APPLICANT: Rhee, Sang-Ki
; APPLICANT: Lee, Eun-Hae
; TITLE OF INVENTION: Gluconobacter Suboxydans Sorbitol Dehydrogenase, Genes
; TITLE OF INVENTION: and Methods of Use Thereof

```



```

; FILE REFERENCE: 1533.0870000
; CURRENT APPLICATION NUMBER: US/09/296,284A
; CURRENT FILING DATE: 1999-04-22
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2265
; TYPE: DNA
; ORGANISM: Gluconobacter suboxydans
; US-09-296-284-1

```

```

Query Match      4.9%; Score 85.6; DB 3; Length 2265;
Best Local Similarity 50.3%; Pred. No. 7, 1e-13;
Matches 304; Conservative 0; Mismatches 279; Indels 21; Gaps 3;

```

```

QY 100 GCGAACCCGCGCGCTGATGATGATTACTACGCGCGCAACCAAGAAAATATGCGCAC 159
DB 130 GCGACACGATCCGGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 189
QY 160 TCGCCCTGACCCAGATCACTGCGCGCAAGTTGATGATGATGATGATGATGATGATGAT 219
DB 190 AGCCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 249
QY 220 GGGATGAGGCGGGGCGCGTACAGTCT--ACGCCATGATCCATGATGATGATGATGATGAT 276
DB 250 GATCTGATACCAACCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 309
QY 277 CTGGCAAAACCCGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 336
DB 310 GCCACCAAACTGGAGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 369
QY 337 GAACACCGCGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 393
DB 370 TCTTACGATCAAAAGTTTCCAGGCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 429
QY 394 CGCGCGCGCGCGCGCTTTCAGGCAAGCTTTCATGATGATGATGATGATGATGATGATGAT 453
DB 430 CGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 489
QY 454 GCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 498
DB 490 GCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 549
QY 499 GGAATGCGGCGCGCGCTTTCAGGCAAGCTTTCATGATGATGATGATGATGATGATGATGAT 558
DB 550 CAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 609
QY 559 ATGCTGCGGCGCGCTTTCAGGCAAGCTTTCATGATGATGATGATGATGATGATGATGAT 618
DB 610 GTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 669
QY 619 TCGCGAGCGGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 678
DB 670 GCTGAAACGGGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 729
QY 679 GACG 682
DB 730 GACG 733

```

```

RESULT 8
US-09-296-284-7
; Sequence 7, Application US/09296284A
; Patent No. 6204040
; GENERAL INFORMATION:
; APPLICANT: Choi, Eun-Sung
; APPLICANT: Rhee, Sang-Ki
; APPLICANT: Lee, Eun-Hae
; TITLE OF INVENTION: Gluconobacter Suboxydans Sorbitol Dehydrogenase, Genes
; FILE REFERENCE: 1533.0870000
; CURRENT APPLICATION NUMBER: US/09/296,284A
; CURRENT FILING DATE: 1999-04-22

```

```

; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 4830
; TYPE: DNA
; ORGANISM: Gluconobacter suboxydans
; US-09-296-284-7

```

```

Query Match      4.9%; Score 85.6; DB 3; Length 4830;
Best Local Similarity 50.3%; Pred. No. 9, 5e-13;
Matches 304; Conservative 0; Mismatches 279; Indels 21; Gaps 3;

```

```

QY 100 GCGAACCCGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 159
DB 794 GCGACACGATCCGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 853
QY 160 TCGCCCTGACCCAGATCACTGCGCGCAAGTTGATGATGATGATGATGATGATGATGATGAT 219
DB 854 AGCCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 913
QY 220 GGGATGAGGCGGGGCGCGTACAGTCT--ACGCCATGATCCATGATGATGATGATGATGAT 276
DB 914 GATCTGATACCAACCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 973
QY 277 CTGGCAAAACCCGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 336
DB 974 GCCACCAAACTGGAGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1033
QY 337 GAACACCGCGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 393
DB 1034 TCTTACGATCAAAAGTTTCCAGGCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1093
QY 394 CGCGCGCGCGCGCGCTTTCAGGCAAGCTTTCATGATGATGATGATGATGATGATGATGAT 453
DB 1094 CGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1153
QY 454 GCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 498
DB 1154 GCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1213
QY 499 GGAATGCGGCGCGCGCTTTCAGGCAAGCTTTCATGATGATGATGATGATGATGATGATGAT 558
DB 1214 CAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1273
QY 559 ATGCTGCGGCGCGCTTTCAGGCAAGCTTTCATGATGATGATGATGATGATGATGATGAT 618
DB 1274 GTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1333
QY 619 TCGCGAGCGGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 678
DB 1334 GCTGAAACGGGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1393
QY 679 GACG 682
DB 1394 GACG 1397

```

```

RESULT 9
US-09-252-991A-15371/c
; Sequence 15371, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubinfeld et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

```

SEQ ID NO 15371  
LENGTH: 1989  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15371

Query Match 4.4%; Score 77; DB 4; Length 1989;  
Best Local Similarity 46.5%; Pred. No. 1.3e-10;  
Matches 248; Conservative 0; Mismatches 285; Indels 0; Gaps 0;

QY 1201 GAGCTGCGCCGACCTTCCTGGTGGGCGGACGCTGCTGACCGGCACTGAACCCGAGC 1260  
DB 696 GAATATGCGCGCGCTTCCTGGGCGGACAGACCTGAACCTGAGCTTACACCCGAGC 637  
QY 1261 ACCGCACTTACTCTTCTGCGCTGAACATGCTGTACGATTTATGCGCTTATCA 1320  
DB 636 ACCGCGCTGTTCTACGTCGCGGCAACCACTGGAAGGAACTACTGACCGAGAGGTC 577  
QY 1321 GAGTTAGCGCGCTGCACTGTATTAACAACAGCGGCAACCTGCGCGGCTTT 1380  
DB 576 AGCTATACGAAGGCGACGCTTACCTTGGCATGGGCTTCGATCAAGCGCATGACAG 517  
QY 1381 GAAATATGCGCGCATGACGCGATGATATGACAGCGGCGCACTTGTGTCGAG 1440  
DB 516 GACACGCTGCGACGCTGCGCGCATGACCGGTCAGCGGCAAGGTGCTGGAAACAC 457  
QY 1441 GAGCGCTTGGCGGAACTACTGCGCTTTTGTGACGCGAGCGGTGTGTTCAAC 1500  
DB 456 AAGGAACACCTGCGCTTGGCGCGGCTGTGCGGACCGCGCAACCTGTGTTCAAC 397  
QY 1501 GCGGAGACCGACCGCTATTTCCGTCGCGCTGACGCGGCAAGAACCGCGGCACTTGTGTCGAG 1560  
DB 396 GGCACCGCGGACCGCTTACTTCAAGGCTTGAAGCGGCAAGAGAGCTGTGGA 337  
QY 1561 GCGGCTTGGCGAGCGGTGCGGACGCGGCGATGACTACGATTGAGCGGCTGCA 1620  
DB 336 TTCAGACCGCGGACCGCATGCTCTGCGCACCATACCTGGGAAACAGACGCGGAGAG 277  
QY 1621 TATATGCGCATGCGTGTGCGGCGGTGACCTATGACGCAATTGAACGCGCGCGGAC 1680  
DB 276 TACCTGCGCGTGAACCGTGTGCGCTACGCGGCGCGCTGCGCGCGCGGACATGAGC 217  
QY 1681 GAGGCATGCACTGCACTGCTGCGGTATGATGATGATGCTTGTGCTGCTGCTG 1733  
DB 216 GACCTGACCGCGCGGCGGCGGCGGCGGCTTCTTGGGTATTCAGCTGCC 164

RESULT 10  
US-09-252-991A-15462  
Sequence 15462, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252.991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074.788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094.190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 15462  
LENGTH: 2058  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15462

Query Match 4.4%; Score 77; DB 4; Length 2058;  
Best Local Similarity 46.5%; Pred. No. 1.3e-10;  
Matches 248; Conservative 0; Mismatches 285; Indels 0; Gaps 0;

QY 1201 GAGCTGCGCCGACCTTCCTGGTGGGCGGACGCTGCTGACCGGCACTGAACCCGAGC 1260  
DB 1495 GAATATGCGCGCGCTTCCTGGGCGGACAGACCTGAACCTGAGCTTACACCCGAGC 1554  
QY 1261 ACCGCACTTACTCTTCTGCGCTGAACATGCTGTACGATTTATGCGCTTATCA 1320  
DB 1555 ACCGCGCTGTTCTACGTCGCGGCAACCACTGGAAGGAACTACTGACCGAGAGGTC 1614  
QY 1321 GAGTTAGCGCGCTGCACTGTATTAACAACAGCGGCAACCTGCGCGGCTTT 1380  
DB 1615 AGCTATACGAAGGCGACGCTTACCTTGGCATGGGCTTCCGATCAAGCCATGATGAC 1674  
QY 1381 GAAATATGCGCGCATGACGCGATGATATGACAGCGGCGCACTTGTGTCGAG 1440  
DB 1675 GACACGCTGCGACCTGCGCGCATGACCGGCTGACGCGGCAAGGTGTGGAACAC 1734  
QY 1441 GAGCGCTTGGCGGAACTACTGCGCGCTTGTGTCGAGCGGCGGTGTGTTCAAC 1500  
DB 1735 AAGGAACACCTGCGCTTGGCGCGGCTGTGCGCACCGCGGCAACCTGTGTTCAAC 1794  
QY 1501 GCGGAGACCGACCGCTATTTCCGTCGCGCTGACCGGAAACCGCGGCACTTGTGTCGAG 1560  
DB 1795 GGCACCGCGGACCGCTTACTTCAAGGCTTGAAGCGGCAAGAGCTGTGGA 1854  
QY 1561 GCGGCTTGGCGAGCGGTGCGGACGCGGCGGATGACTACGAGTTGAGCGGCTGCA 1620  
DB 1855 TTCAGACCGGACGCGCATGCTGCGCACCATCTGAGGAAACAGACGCGGAGAG 1914  
QY 1621 TATATGCGCATGCGTGTGCGGCGGTGACCTATGACGCAATTGAACGCGCGCTGAGC 1680  
DB 1915 TACCTGCGCGTGAACCGTGTGCGGTACGCGGCGCGCTGCGTGGGCGGCGGACATGAGC 1974  
QY 1681 GAGGCATGCACTGCACTGCTGCGGTATGATGATGATGCTTGTGCTGCTGCTG 1733  
DB 1975 GACCTGACCGCGCGGCGGCGGCGGCGGCTTCTTGGGTATTCAGCTGCC 2027

RESULT 11  
US-09-252-991A-15403  
Sequence 15403, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252.991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074.788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094.190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 15403  
LENGTH: 2637  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15403

Query Match 4.4%; Score 77; DB 4; Length 2637;  
Best Local Similarity 46.5%; Pred. No. 1.4e-10;  
Matches 248; Conservative 0; Mismatches 285; Indels 0; Gaps 0;

QY 1201 GAGCTGCGCCGACCTTCCTGGTGGGCGGACGCTGCTGACCGGCACTGAACCCGAGC 1260  
DB 1373 GAATATGCGCGCGCTTCCTGGGCGGACAGACCTGAACCTGAGCTTACACCGAGAGC 1432  
QY 1261 ACCGCACTTACTCTTCTGCGCTGAACATGCTGTACGATTTATGCGCTTATCA 1320  
DB 1433 ACCGCGCTGTTCTACGTCGCGGCAACCACTGGAAGGAACTACTGACCGAGAGGTC 1492  
QY 1321 GAGTTAGCGCGCTGCACTGTATTAACAACAGCGGCAACCTGCGCGGCTTT 1380

Db 1493 AGCTATACGAGGAGCGGCTTCTTGGCATGAGCTTCCGATCAAGCGATGATGAC 1552  
Qy 1381 GAAATATGAGCGCGCATCGAGATTGATATGACGAGCGGCGCATCTTGATGATGAG 1440  
Db 1553 GACCACTGCGGAGCTTGGCGGATGAGCGGCGGATGAGCGGCGGATGAGCGG 1612  
Qy 1441 GAGCGCTGCGGCGGATGAGCTTGGCGGATGAGCGGCGGATGAGCGGCGG 1500  
Db 1613 AAGGAAACACCTGCGGCTTGGCGGATGAGCGGCGGATGAGCGGCGG 1672  
Qy 1501 GAGCGGAGCGGCGGATGAGCTTGGCGGATGAGCGGCGGATGAGCGGCGG 1560  
Db 1673 GAGCGGAGCGGCGGATGAGCTTGGCGGATGAGCGGCGGATGAGCGGCGG 1732  
Qy 1561 GCGCGCTTGGCGGATGAGCTTGGCGGATGAGCGGCGGATGAGCGGCGG 1620  
Db 1733 TTCGAGCGGCGGATGAGCTTGGCGGATGAGCGGCGGATGAGCGGCGG 1792  
Qy 1621 TATATGCGGATGAGCTTGGCGGATGAGCGGCGGATGAGCGGCGG 1680  
Db 1793 TACCTGCGGATGAGCTTGGCGGATGAGCGGCGGATGAGCGGCGG 1852  
Qy 1681 GAGCGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCGGCGG 1733  
Db 1853 GACCTGAGCGGCGGATGAGCTTGGCGGATGAGCGGCGG 1905

RESULT 12  
US-09-252-991A-15512  
; Sequence 15512, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-07-27  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 15512  
; LENGTH: 510  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15512

Query Match 4.2%; Score 72.6; DB 4; Length 510;  
Best Local Similarity 48.0%; Pred. No. 1.1e-09;  
Matches 207; Conservativity 0; Mismatches 224; Indels 0; Gaps 0;  
Qy 1201 GAGCTGCGGCGGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGG 1260  
Db 78 GAGGATGCGGCGGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGG 137  
Qy 1261 ACCGCGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGG 1320  
Db 138 ACCGCGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGG 1397  
Qy 1321 GAGGCGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGG 1380  
Db 198 AGGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGG 257  
Qy 1381 GAAATATGAGCGCGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGG 1440  
Db 258 GACCGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGG 317  
Qy 1441 GAGCGCGGCGGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGG 1500  
Db 318 AAGGAAACACCTGCGGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGG 377

Qy 1501 GCGGAGCGGCGGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGG 1560  
Db 378 GCGGAGCGGCGGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGG 437  
Qy 1561 GCGGCGGCGGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGG 1620  
Db 438 TTCGAGCGGCGGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGG 497  
Qy 1621 TATATGCGGAT 1631  
Db 498 TACCTGCGGCT 508

RESULT 13  
US-09-489-039A-5265  
; Sequence 5265, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 5265  
; LENGTH: 1809  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-5265

Query Match 3.1%; Score 53.8; DB 4; Length 1809;  
Best Local Similarity 60.5%; Pred. No. 0.00016;  
Matches 107; Conservativity 0; Mismatches 67; Indels 3; Gaps 1;  
Qy 1472 TGTCAGCGGCGGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGG 1528  
Db 1565 TCTCAGCGGCGGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGG 1624  
Qy 1529 TCAGCGGAGGAGCGGCGGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGG 1588  
Db 1625 ACAATATGAGCGGCGGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGG 1684  
Qy 1589 AGCGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGG 1645  
Db 1685 CTCGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGGATGAGCTTGGCGG 1741

RESULT 14  
US-08-476-519-10  
; Sequence 10, Application US/08476519  
; Patent No. 5750876  
; GENERAL INFORMATION:  
; APPLICANT: Barry, Gerard F.  
; APPLICANT: Kishore, Ganesh M.  
; TITLE OF INVENTION: No. 5750876el Isoamylase Gene, Compositions  
; TITLE OF INVENTION: Containing It and Methods of Using Isoamylases  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESSES:  
; ADDRESS: Grace L. Bonner, Monsanto Company, B44F  
; STREET: 700 Chesterfield Parkway No. 5750876th  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA  
; ZIP: 63198  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/476,519
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/281902
FILING DATE: 28-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bonner, Grace L.
REGISTRATION NUMBER: 32,963
REFERENCE/DOCKET NUMBER: 38-21(13577)A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314)537-7286
TELEFAX: (314)537-6047
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 224 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2241
US-08-476-519-10

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: November 22, 2004, 14:45:05 ; Search time 933 Seconds

(without alignments)  
10077.497 Million cell updates/sec

Title: US-10-802-682-4

Perfect score: 1 atgacccccacaacgctgct.....ttcttgacactgcgcagctaa 1740

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3627888 seqs, 2701811610 residues

Total number of hits satisfying chosen parameters: 7255776

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*  
1: /cgn2\_6/prodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/prodata/1/pubpna/PCR\_NEW\_PUB.seq:\*  
3: /cgn2\_6/prodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/prodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
5: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq:\*  
6: /cgn2\_6/prodata/1/pubpna/PCRUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/prodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/prodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq:\*  
12: /cgn2\_6/prodata/1/pubpna/US09\_PUB.seq:\*  
13: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq:\*  
15: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq:\*  
16: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq:\*  
17: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq:\*  
18: /cgn2\_6/prodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
19: /cgn2\_6/prodata/1/pubpna/US11\_NEW\_PUB.seq:\*  
20: /cgn2\_6/prodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
21: /cgn2\_6/prodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1206.2	69.3	1743	US-10-162-713-4	Sequence 4, Appl1
2	1206.2	69.3	2281	US-10-162-713-11	Sequence 11, Appl1
3	1071.2	61.6	1737	US-10-162-713-1	Sequence 1, Appl1
4	1071.2	61.6	2519	US-10-162-713-9	Sequence 9, Appl1
5	1052	60.5	1740	US-10-162-713-3	Sequence 3, Appl1
6	1052	60.5	3200	US-10-162-713-10	Sequence 10, Appl1
7	152.2	8.7	2151	US-10-369-493-34527	Sequence 34527, A
8	99.4	5.7	1731	US-10-369-493-45426	Sequence 45426, A
9	83.8	4.8	1677	US-10-369-493-28294	Sequence 28294, A
10	83.8	4.8	1701	US-10-369-493-31051	Sequence 31051, A
11	76.6	4.4	2076	US-10-369-493-44211	Sequence 44211, A
12	74.8	4.3	1596	US-10-369-493-31374	Sequence 31374, A

13	74.8	4.3	2051	15	US-10-369-493-28615	Sequence 28616, A
14	74	4.3	1869	15	US-10-369-493-31733	Sequence 31733, A
15	68.8	4.0	2120	15	US-10-369-493-34525	Sequence 34525, A
16	65.8	3.8	2412	15	US-10-369-493-28508	Sequence 28508, A
17	65.8	3.8	2412	15	US-10-369-493-31267	Sequence 31267, A
18	62.8	3.6	1803	15	US-10-369-493-31375	Sequence 31375, A
19	62.8	3.6	1803	15	US-10-369-493-28617	Sequence 28617, A
20	61	3.5	2331	15	US-10-369-493-31725	Sequence 31725, A
21	57	3.3	2376	15	US-10-369-493-28565	Sequence 28565, A
22	57	3.3	2376	15	US-10-369-493-31324	Sequence 31324, A
23	53.6	3.1	2394	15	US-10-369-493-39100	Sequence 39100, A
24	53.6	3.1	2394	15	US-10-369-493-39468	Sequence 39468, A
25	53.6	3.1	2394	15	US-10-369-493-39850	Sequence 39850, A
26	53.6	3.1	4412	10	US-09-927-827-43	Sequence 43, Appl1
27	50	2.9	1776	15	US-10-369-493-39656	Sequence 39656, A
28	50	2.9	1776	15	US-10-369-493-40018	Sequence 40018, A
29	50	2.9	1779	15	US-10-369-493-39280	Sequence 39280, A
30	49.4	2.8	2400	15	US-10-369-493-37751	Sequence 37751, A
31	47	2.7	7155	15	US-10-329-079-14	Sequence 14, Appl1
32	47	2.7	37360	15	US-10-329-079-6	Sequence 6, Appl1
33	46.6	2.7	536	16	US-10-338-110-119	Sequence 119, App
34	46	2.6	2391	15	US-10-369-493-24390	Sequence 24390, A
35	45.8	2.6	1401	11	US-09-758-759-106	Sequence 106, App
36	45.8	2.6	109519	11	US-09-758-759-1	Sequence 1, Appl1
37	45	2.6	2358	15	US-10-369-493-35570	Sequence 35570, A
38	44.6	2.6	1899	15	US-10-156-761-5926	Sequence 5926, Ap
39	44.6	2.6	9025608	15	US-10-156-761-1	Sequence 1, Appl1
40	44.4	2.6	851	18	US-10-425-115-12182	Sequence 12182, A
41	44.4	2.6	893	18	US-10-425-115-12183	Sequence 12183, A
42	44.2	2.5	1860	15	US-10-369-493-40752	Sequence 40752, A
43	43.6	2.5	894	16	US-10-282-1224-25434	Sequence 25434, A
44	43.6	2.5	1338	15	US-10-369-493-31938	Sequence 31938, A
45	43.4	2.5	1049	14	US-10-123-155-358	Sequence 358, App

#### ALIGNMENTS

RESULT 1  
US-10-162-713-4  
Sequence 4, Application US/10162713  
Publication No. US20030228672A1  
GENERAL INFORMATION:  
APPLICANT: Choi, Eun-Sung  
APPLICANT: D'Elia, John  
APPLICANT: Kim, Hye-Sun  
APPLICANT: Kim, Mi-Soo  
APPLICANT: Lee, Jung Kee  
APPLICANT: Pan, Jae-Gu  
APPLICANT: Stoddard, Steven F.  
TITLE OF INVENTION: Sorbital Dehydrogenases of Ketogulonigenium spp., Genes and Methoc  
FILE REFERENCE: 1533.3740000  
CURRENT APPLICATION NUMBER: US/10/162,713  
CURRENT FILING DATE: 2002-06-06  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 4  
LENGTH: 1743  
TYPE: DNA  
ORGANISM: Ketogulonigenium sp.  
US-10-162-713-4  
Query Match 69.3%; Score 1206.2; DB 15; Length 1743;  
Best Local Similarity 80.9%; Pred. No. 0;  
Matches 1406; Conservative 0; Mismatches 333; Indels 0; Gaps 0;  
QY 1 ATGACCCCAACACGCTGCTTGCACACGCGGCGGCGCTATGCTTACCGCGCCGCC 60  
DB 1 ATGCGACCCCAACACGCTGCTTGCACACGCGGCGGCGCTATGCTTACCGCGCCGCC 60  
QY 61 GCATTCGCGCAGTAACCCGATTAACGATGAATCTGCGGACCCGCCCGCTGTGTA 120

Db 61 GCGTTTCGAGGCTGACCCCATACCGATGAACTGTGCGCAACCGGCGACGGCGGAG 120  
Qy 121 TGAATTAACTACGCGCGGCAACCAAGAAATATCGCACTCGGCGCTGACCGCATCACT 180  
Db 121 TGAATTAACTACGCGCGGCAACCAAGAAATATCGCACTCGGCGCTGACCGCATCACT 180  
Qy 181 GCCGCAACAGTGTGCTAGTTGCAACTGCTGTGGGCGCGCGGAGTGAAGCGGGGCGCTTA 240  
Db 181 ACCGACAACGTCGCGCGAGCTGAGCTGTGTGGGCGCGCGGAGTGAAGCGGGGCGCTTA 240  
Qy 241 CAGGTCAACGCGCGATGATCACTGATGAGGTGATGATGATGCAAAACCGCGTATGATG 300  
Db 241 CAGGTCAACGCGCGATGATCACTGATGAGGTGATGATGATGCAAAACCGCGTATGATG 300  
Qy 301 CAGGCGCTGATGCGCAACAGCGGATCTGATCTGGGCAACACCGCGCGCAACTGCGCGC 360  
Db 301 CAGGCGCTGATGCGCAACAGCGGATCTGATCTGGGCAACACCGCGCGCAACTGCGCGC 360  
Qy 361 GTGCGCAAGCTTAAACGCGCAAGCGGACCGGCGCGCGCTGCGCTTTACGCGCACGAC 420  
Db 361 GTGCGCAAGCTTAAACGCGCAAGCGGACCGGCGCGCGCTGCGCTTTACGCGCACGAC 420  
Qy 421 CTGATTTGAGCTCATGAGGCAACCACTGATGCGCGTGGATATGAGACGCGGCGAGTC 480  
Db 421 CTGATTTGAGCTCATGAGGCAACCACTGATGCGCGTGGATATGAGACGCGGCGAGTC 480  
Qy 481 GATTCGATGTCGACGCTGATGCGGCGGAGGAGCGGCTTGAACGATTAACCAACGCGGCG 540  
Db 481 GATTCGATGTCGACGCTGATGCGGCGGAGGAGCGGCTTGAACGATTAACCAACGCGGCG 540  
Qy 541 ATGTGCGCAATGCGCGATGCTGCGGCGGTCGACCTGCGCAATATGCGCGATGATGATG 600  
Db 541 ATGTGCGCAATGCGCGATGCTGCGGCGGTCGACCTGCGCAATATGCGCGATGATGATG 600  
Qy 601 TTTATCTCGGCGGACGATTCGCGGAGCGGAGCTGTGCGCAACCACTTTATCCG 660  
Db 601 TTTATCTCGGCGGACGATTCGCGGAGCGGAGCTGTGCGCAACCACTTTATCCG 660  
Qy 661 CAGCCGCGGCAAAAGGCTGACGAGACTTGGGCGCAATGATTTGAGGCGCGCTGATGAC 720  
Db 661 CAGCCGCGGCAAAAGGCTGACGAGACTTGGGCGCAATGATTTGAGGCGCGCTGATGAC 720  
Qy 721 GCGGCTGCGGCGTCAATCACTTATGATCCGCTGACGCAACTTGTGTCTATGCTGAC 780  
Db 721 GCGGCTGCGGCGTCAATCACTTATGATCCGCTGACGCAACTTGTGTCTATGCTGAC 780  
Qy 781 GCGGCTGCGGCGTCAATCACTTATGATCCGCTGACGCAACTTGTGTCTATGCTGAC 840  
Db 781 GCGGCTGCGGCGTCAATCACTTATGATCCGCTGACGCAACTTGTGTCTATGCTGAC 840  
Qy 841 AACACCGGCTTTGCGGCGTCCGCAACGCGGCGGAGATGCTGTGCGGTGACGACCGCTG 900  
Db 841 AACACCGGCTTTGCGGCGTCCGCAACGCGGCGGAGATGCTGTGCGGTGACGACCGCTG 900  
Qy 901 CCGCGCGCAACTGCGGCAAGAGTCACTGATGATGATGATGATGATGATGATGATGATG 960  
Db 901 CCGCGCGCAACTGCGGCAAGAGTCACTGATGATGATGATGATGATGATGATGATGATG 960  
Qy 961 CAGGCGCTGCGGCAATGAGCGGCTGCGCGGCGATTAACCGCAATGCGGCGGCGGAG 1020  
Db 961 CAGGCGCTGCGGCAATGAGCGGCTGCGCGGCGATTAACCGCAATGCGGCGGCGGAG 1020  
Qy 1021 CCGCGCTGCGGCAATGAGCGGCTGCGCGGCGATTAACCGCAATGCGGCGGCGGAG 1080  
Db 1021 CCGCGCTGCGGCAATGAGCGGCTGCGCGGCGATTAACCGCAATGCGGCGGCGGAG 1080  
Qy 1081 TCGCGGCAATGCTGCGGCGGCGGCTGATACCAACTATACCAACTATGATGCGCTGATGAC 1140  
Db 1081 ACCGCGCAATGCTGCGGCGGCGGCTGATACCAACTATACCAACTATGATGCGCTGATGAC 1140  
Qy 1141 GAGACCGGCTTGTGACGCTGAGACGAGATGCGGTGCTGAAAGAGCTGAGCGTTGATAT 1200  
Db 1141 GAGACCGGCTTGTGACGCTGAGACGAGATGCGGTGCTGAAAGAGCTGAGCGTTGATAT 1200

Db 1141 GAAACCGGCTGTCACGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
Qy 1201 GAGGCTGCGGAGCGCTTCTGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
Db 1201 GAGGCTGCGGAGCGCTTCTGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
Qy 1261 ACCGCGCAATGCTTCTGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
Db 1261 ACCGCGCAATGCTTCTGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
Qy 1321 GAGGCTGCGGAGCGCTTCTGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
Db 1321 GAGGCTGCGGAGCGCTTCTGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
Qy 1381 GAAATATGAGGCGGAG 1440  
Db 1381 GAAATATGAGGCGGAG 1440  
Qy 1441 GAGGCGCTTCTGCGGAG 1500  
Db 1441 GAGGCGCTTCTGCGGAG 1500  
Qy 1501 GCGGCGAGCGGAG 1560  
Db 1501 GCGGCGAGCGGAG 1560  
Qy 1561 GCGGCTTCTGCGGAG 1620  
Db 1561 GCGGCTTCTGCGGAG 1620  
Qy 1621 TATATGCGCAATGCGGAG 1680  
Db 1621 TATATGCGCAATGCGGAG 1680  
Qy 1681 GAGGCAATGATGATGAGCTGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740  
Db 1681 GAGGCAATGATGATGAGCTGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740

RESULT 2  
US-10-162-713-11  
; Sequence 11, Application US/10162713  
; Publication No. US20030228672A1  
; GENERAL INFORMATION:  
; APPLICANT: Choi, Eun-Sung  
; APPLICANT: D'Elia, John  
; APPLICANT: Kim, Hye-Sun  
; APPLICANT: Kim, Mi-Soe  
; APPLICANT: Lee, Jung Kee  
; APPLICANT: Lee, Jae-Gu  
; APPLICANT: Stoddard, Steven F.  
; TITLE OF INVENTION: Sorbitol Dehydrogenases of Ketogulonigenium spp., Genes and Method  
; TITLE OF INVENTION: Use thereof  
; FILE REFERENCE: 1533.3740000  
; CURRENT APPLICATION NUMBER: US/10/162, 713  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 11  
; LENGTH: 2281  
; TYPE: DNA  
; ORGANISM: Ketogulonigenium sp.  
US-10-162-713-11

Query Match 69.3%; Score 1206.2; DB 15; Length 2281;  
Best Local Similarity 80.9%; Pred. No. 0;  
Matches 1406; Conservative 0; Mismatches 333; Indels 0; Gaps 0;

Qy 1 ATGAAACCCCAACAGCTGCTTTCGACCAAGCGCGCGCTGCTATGCTTACCGCGCGCC 60  
Db 530 ATGGAACCCCAACAGCTGCTTTCGACCAAGCGCGCGCTGCTATGCTTACCGCGCGCC 589

QY 61 GCATTGCGGAGGTAAACCCGATTAACCGATGAAGTCTGGCGAAACCCGCCGCTGTGAA 120  
 Db 590 GCGTTGGCGGAGGTACACCCCATACCGATGAAGTCTGGCGAAACCCGCCGCGGCGAG 649  
 QY 121 TGGATTAACTACGGCGGCAACCAAGAAAATCTACCGCACTCGCCCTTGAACCAATCACT 180  
 Db 650 TGGATTAACTACGGCGGCAATCAGAAAATCTACCGCACTCGCCCTTGAACCAATCACT 709  
 QY 181 GCGCAAGTGTGTGAGTGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 240  
 Db 710 ACCGACAAAGT 769  
 QY 241 CAGGTACACCGCGATGATTCATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
 Db 770 CAGGTACACCGCGATGATTCATGATGATGATGATGATGATGATGATGATGATGATGAT 829  
 QY 301 CAGGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360  
 Db 830 CAGGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 889  
 QY 361 GTCGCAACGCTAAACCGCAAGCGCGACCGCAAGCGCGCGCTTACCGCAACGAC 420  
 Db 890 GTTGCTCGCTGACGCGCGCAAGCGCGACCGCAAGCGCGCGCTTACCGCAACGAC 949  
 QY 421 CTCTATTTCAGCTCATGAGCAACCATCTGATGCGCTGATGATGATGATGATGATGATGAT 480  
 Db 950 CTCTATTTCAGCTCATGAGCAACCATCTGATGCGCTGATGATGATGATGATGATGATGAT 1009  
 QY 481 GTTATGATGTGAAAGT 540  
 Db 1010 GTTATGATGTGAAAGT 1069  
 QY 541 ATTGTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
 Db 1070 ATTGTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1129  
 QY 601 TTATATCTGGGCGGACGATTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 660  
 Db 1130 TTATATCTGGGCGGACGATTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1189  
 QY 661 CAGCGCGGCGAAGAGGAGTGAAGAGCTTGGGCGAATGATTCGAGGCGGCGTGTGATGAC 720  
 Db 1190 CAGCGCGGCGAAGAGGAGTGAAGAGCTTGGGCGAATGATTCGAGGCGGCGTGTGATGAC 1249  
 QY 721 GCGGCTCTGGGAGTGAATCACTTATGATCCCGTGAACCACTTGTGTGTGTGTGTGTGTGT 780  
 Db 1250 GCGGCTCTGGGAGTGAATCACTTATGATCCCGTGAACCACTTGTGTGTGTGTGTGTGT 1309  
 QY 781 GCGGCGGCGCGACGTCGCGAACCAGCGCGGCGACCGCGGCGGCGGCGGCGGCGGCGGCG 840  
 Db 1310 GCGGCGGCGCGACGTCGCGAACCAGCGCGGCGACCGCGGCGGCGGCGGCGGCGGCGGCG 1369  
 QY 841 AACACCGGCTTGGCGGTCGTCGCGAAGCGGCGAATGTCGTCGTCGTCGTCGTCGTCGTCG 900  
 Db 1370 AACACCGGCTTGGCGGTCGTCGCGAAGCGGCGAATGTCGTCGTCGTCGTCGTCGTCGTCG 1429  
 QY 901 CCGCGCGAAGTCTGGGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATG 960  
 Db 1430 CCGCGCGAAGTCTGGGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1489  
 QY 961 CAAACCGCTCGGCGAGATGAGAGGTCTGCGCGCATCAACCCCAATGCGGCGAGCGGCGAG 1020  
 Db 1490 CAAACCGCTCGGCGAGATGAGAGGTCTGCGCGCATCAACCCCAATGCGGCGAGCGGCGAG 1549  
 QY 1021 CGCGGTGTGTGACGAGGTGCGCTTGAAGACCGGCGACATGTGTGTGTGTGTGTGTGTGTGT 1080  
 Db 1550 CGCGGTGTGTGACGAGGTGCGCTTGAAGACCGGCGACATGTGTGTGTGTGTGTGTGTGTGT 1609  
 QY 1081 TCGGCGCAATTCCTGTGGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
 Db 1610 TCGGCGCAATTCCTGTGGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1669  
 QY 1141 GAGACCGGCTTGTGACGAGTGAACGAGAGATGCGGTGTGAAGAGCTGACGATGTGATAT 1200

Db 1670 GAAACCGGCTGTGACAGTGAAGCAAGAGAGGTGTGTGACGAACTGTGACACCGATAC 1729  
 QY 1201 GAGCTCTGCGCGACCTTCTCGGGGTGGGCGGACGTCGTGACAGCGGCGATGAACCGGAC 1260  
 Db 1730 GAGCTCTGCGCGACCTTCTCGGGGTGGGCGGACGTCGTGACAGCGGCGATGAACCGGAT 1789  
 QY 1261 ACCGCAATTTACTTCTTCCGCTGAAACATGCTGTCTACGATATATGAGCGGTGATCA 1320  
 Db 1790 AGCGCAATTTACTTCTTCCGCTGAAACATGCTGTCTACGATATATGAGCGGTGATCA 1849  
 QY 1321 GAGTTAGCGGCTGACGTCATTAACAGAGCGGACCGCAAAATCTGGGCGGCGGCTT 1380  
 Db 1850 GAGTTAGCGGCTGACGTCATTAACAGAGCGGACCGCAAAATCTGGGCGGCGGCTT 1909  
 QY 1381 GAAATATGAGCGGCTGACGTCATTAACAGAGCGGACCGCAAAATCTGGGCGGCGGCTT 1440  
 Db 1910 GAAATATGAGCGGCTGACGTCATTAACAGAGCGGACCGCAAAATCTGGGCGGCGGCTT 1969  
 QY 1441 GAGCGCGCTGTGCGGAACTACTGCGCCGTTTGTGTGACGCGAGCGGTGTGTGTGAC 1500  
 Db 1970 GAGCGCGCTGTGCGGAACTACTGCGCCGTTTGTGTGACGCGAGCGGTGTGTGTGAC 2029  
 QY 1501 GCGCGGACCGGACGCTATTTCCGTCGCTCAGACGAGGAAACCGGCGAGACTTGTGTGAC 1560  
 Db 2030 GCGCGGACCGGACGCTATTTCCGTCGCTCAGACGAGGAAACCGGCGAGACTTGTGTGAC 2089  
 QY 1561 GCGCGCTTGTGCGAGTGTGCGACGAGGCGAGCGATGACGTAAGATTGACAGCGGTGCA 1620  
 Db 2090 ACCGCTGTGGGAGTGTGCGTACCGGCGAAGCGCATGAGTAAATGACAGCGGACCA 2149  
 QY 1621 TATATGCGCATGCGGTGGGCGGCTGTGACCTATGCGAGCGCATGATGAGCGGCGCTGAC 1680  
 Db 2150 TATATGCGCATGCGGTGGGCGGCTGTGACCTATGCGAGCGCATGATGAGCGGCGCTGAC 2209  
 QY 1681 GAGCAATGCAATTCGACCTCGGTGTGATGCGATGATGATGATGATGATGATGATGATGAT 1739  
 Db 2210 GAGCAATGCAATTCGACCTCGGTGTGATGCGATGATGATGATGATGATGATGATGATGAT 2268

RESULT 3  
 US-10-162-713-1  
 ; Sequence 1, Application US/10162713  
 ; Publication No. US20030228672A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Choi, Eun-Sung  
 ; APPLICANT: D'Elia, John  
 ; APPLICANT: Kim, Hye-Sun  
 ; APPLICANT: Kim, Mi-Soo  
 ; APPLICANT: Lee, Jung Kee  
 ; APPLICANT: Pan, Jae-Gu  
 ; APPLICANT: Stoddard, Steven F.  
 ; APPLICANT: Yum, De-Young  
 ; TITLE OF INVENTION: Sorbitol Dehydrogenases of Ketogulonigenium spp., Genes and Method  
 ; FILE REFERENCE: Use Thereof  
 ; CURRENT APPLICATION NUMBER: 1533.374000  
 ; CURRENT FILING DATE: 2002-06-06  
 ; NUMBER OF SEQ ID NOS: 25  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 1  
 ; LENGTH: 1737  
 ; TYPE: DNA  
 ; ORGANISM: Ketogulonigenium sp.  
 US-10-162-713-1

Query Match 61.6%; Score 1071.2; DB 15; Length 1737;  
 Best Local Similarity 76.4%; Fred. No. 1.2e-308;  
 Matches 1329; Conservative 0; Mismatches 408; Indels 3; Gaps 1;  
 QY 1 ATGAACCCCAAGACGCTGCTTGTGACACGCGCGGCGGCTATGCTTACCGCGCCGCC 60  
 Db 1 ATGAATGGAATTCGTTCTTGTGCGAAGCGTGTGCGGCTTGTGATTCCTTGTGCGCC 60

```

Qy 61 GCATTGCGGAGGTAAACCCGATTAACGATGAATGCTGGCGAACCCGCCGCTGGTGA 120
Db 61 GCAATTTGCGGATGAGAGCCCGTCAACGAGCTGTATGAAACCCGCCGCGCGGAA 120
Qy 121 TGGATTACTACGGCGGCAACAAGAAAACTATCGCACTCGCCCGTGAACCGATCACT 180
Db 121 TGGATTAGCTATGGCGGCAACAAGAAAACTATCGCACTCGCCCGTGAACCAATCAAC 180
Qy 181 GCCGCAACGTTGGTCACTGGTCTGGGAGCCCGCGGATGAGAGCGGAGCGGTA 240
Db 181 CCGGCAACGTTGGGCGGCTGCGAGCTGGTCTGGGAGCGCGGATGAGAGCCCGCGTCTG 240
Qy 241 CAGGTCAAGCGGATGATCACTGATGGGTGATGATATCTGGCAACCCCGGTGATGATC 300
Db 241 CAGGTCAAGCGGCTGATCACTGAGCGGTGATGATCTGGCAACCCAGCGATCACTATT 300
Qy 301 CAGCGCTGAGATGCGCAACAGCGGATCTGATCTGGGAAACCGCGCGCACTGCCCGC 360
Db 301 CAGCGGATGAGCGCAACCGGATCTGATCTGGGAAACCGCGCGCACTGCCCGAG 360
Qy 361 GTGGCGACGCTAAACCGCGGAGCGGAGCGCGGCGCTGGCTTTACGGCAAGAC 420
Db 361 ACCTGACGCTGAGCTGCTGGGAGATCGCAACGCGGCGATCGCGCTTTATGGCACAAT 420
Qy 421 CTCTATTGAGCTCATGAGCAACAACCTGATCGCTGATGATGAGAGCGGCGCAGTTC 480
Db 421 GTCTACTTGTCTGTGGGCAACAACATGATCGCGCTGATGATCTGCCAGCGGCAAGTC 480
Qy 481 GTATTGATGATGAGTGTGATCGGCGGAGAGCGGCTTGAACGATTAACACAGCGGCGC 540
Db 481 GTCTTGGACGCTCAACCGCGGCGCAAGCGGAGAGCGGAGCTGCG--AACTGTCGCGCC 537
Qy 541 ATTGTGCGCAATGAGCGCTGATCGCGGAGTTCACCTGCGCAATATTCGCGCTATGATGC 600
Db 538 ATTGTGCGCAACGCGCTGATCGTGCAGCGGTTGACCTGCGCAATATTCGCGCTTCCGCTGT 597
Qy 601 TTTATCTCGGCGGCAAGATTCGCGAGCGGAGTGAAGACCTGTGGCGCAACCACTTTATCCG 660
Db 598 TTTGTGTGCGGCGCATATGCAACGAGCGGAGAACTGTGGCGCAACTCACTTCACTCCG 657
Qy 661 CAGCGGCGGAGAGGAGTGAACGAGCTGGGCAATGATTTGAGGCGGCGTGGATGAC 720
Db 658 CAGGCAAGTAAAGAGGTGACGAAACCTGGGCAATGATTAAGAAACCGCGCTGATGAC 717
Qy 721 GAGGCTGAGGAGTGAATCACTATGATCCGTGAGCAACCTTGTGTCTATGAGCTGAC 780
Db 718 GAGGCTGAGGAGTGAATCACTATGACCCCACTAATTTGTGTATTAAGGATCGTCG 777
Qy 781 GAGGCTGAGGAGTGAATCACTATGATCCGTGAGCAACCTTGTGTCTATGAGCTGAC 840
Db 778 GAGGCTGAGGAGTGAATCACTATGATCCGTGAGCAACCTTGTGTCTATGAGCTGAC 837
Qy 841 AACACCGGCTTTGCGGTGCGTCCCGACAGCGGCGAGATTTGTGTGCGTCAACGACCTG 900
Db 838 AACACCGGCTTTGCGGTGCGTCCCGACAGCGGCGAGATTTGTGTGCGTCAACGACCTG 897
Qy 901 CCGCGGAGCAACTGAGCAACAAGATGCAAGTGTGAGATGATGTCGCCCACTGATGTC 960
Db 898 CCGCGGAGCAACTGAGCAACAAGTGTGCAAGTGTGAGATGATGTCGCCCACTGATGTC 957
Qy 961 CAACCTGCGGCGAGATGAGAGGAGTCTGGCGGCGCATCAACCCCAATTCGCGGAGCGAG 1020
Db 958 CAGCGCGGCTGCGCACTGAGAGCGGCGTGAAGCGCATCAACCCCAATTCGCGGAGCGAG 1017
Qy 1021 CCGCGTGTGCTGAAGGAGTGGCTTGAAGACCGGCAAGATGATGTCGTTGATGCGGC 1080
Db 1018 CCGTGTGCTGAGCGGCGTTCGCTGCAAAACCGGATGATGTCGAGTTCAGCCCTGA 1077
Qy 1081 TCGGCGGAATTCCTGTGGGCGGTGATACCAACTATACCAATATGATGCGCTCGATGAC 1140
Db 1078 AGGAGGAAATTCCTGTGGGCGGTGATACCAACTATACCAATATGATGCTGATGAC 1137

```

```

Qy 1141 GAGACCGGCTTTGATGAGGTGAACGAGATCGGCTGTGAAGAGCTGACGTTGAATAT 1200
Db 1138 GAAACCGGCTTTGATGAGGTGAATGAATATCACTCTAAAGATCTGACACCGACTAC 1197
Qy 1201 GACGCTGCGGCACTTCTGAGTGGCGGAGACTGTGTCAGCGGCACTGAACCCCGAC 1260
Db 1198 GCGATTTGCGGCACTTCTGAGTGGCGGAGACTGTGTCAGCGGCACTGAACCCCGAT 1257
Qy 1261 ACCGCAATTACTTCTTCCGCTGAACAAATGCTGCTACATATATATAGCGGCTTGA 1320
Db 1258 AGCGCACTACTTCTTCTTCCGCTGAACAAACGCTGCGGATTTGGCGGAGTGCATCAA 1317
Qy 1321 GAGTTAGCGGCTGAGAGTATATACAGCGGCAACCCCAAACTCGCGCGGCGCTT 1380
Db 1318 GAGTTACGCGGAATGAGAGTATACAGCGGCAACCCCAAACTCGCGCGGCGGAAAAA 1377
Qy 1381 GAAATATGCGCGCATGAGCGGATGATATACAGCGGCGGCACTTGTGTGCGG 1440
Db 1378 GAAACATGCGGCGGCGCATGAGCGGATGATATACAGCGGCGGCAACCCGTGTGTGTC 1437
Qy 1441 GAGGCGGCTGCGGGAATCACTGCGCGCTTTGTGAGCGGCGGCGGCTGTGTGCAAC 1500
Db 1438 GAGGCTGCGGCTGAGAGTACTGCGCGCTCTCTGAGCGGCTGCGGCGGCTGTGTGCAAC 1497
Qy 1501 GCGGAGACCGACCGCTATTTCCGTGCGGCTGAGCGGCAACCGGCGAGCTTTGTGCGAG 1560
Db 1498 GCGGCGAGCGATCTGATCTCCGTCCTGAGCGGAACTGCGGAGACCTGTGCGAG 1557
Qy 1561 GCGGCTTGTGAGAGTGTGCGAGCGGCGGAGCGATGAGTATGAGTGTGAGCGGCGGCA 1620
Db 1558 ACCGCTGCGGAGTGTGCGAGCGGCGGAGCGATGAGTATGAGTGTGAGCGGCGGCGAG 1617
Qy 1621 TATATGCGCATGCGGTGCGGCGGTGAGCTATGAGCGGCAATGAAACCGCGCGTGGCC 1680
Db 1618 TATGTGCGCATGCGGAGCGGCGGTATATCTGAGCGGCAATGAAACGAGCAATTCGCG 1677
Qy 1681 GAGGCAATGAGTGAAGTCTGCTGTGTGATGAGTATGATCTTGTGACCTGCGGAGTAA 1740
Db 1678 GCGACCATGAGTGTGAGTGTGATGAGCGGCGGTATGATGCTTGTGCGGCAATTA 1737

```

```

RESULT 4
US-10-162-713-9
; Sequence 9, Application US/10162713
; Publication No. US20030228672A1
; GENERAL INFORMATION:
; APPLICANT: Choi, Eui-Sung
; APPLICANT: D'Elia, John
; APPLICANT: Kim, Hye-Sun
; APPLICANT: Kim, Mi-Soo
; APPLICANT: Lee, Jung Kee
; APPLICANT: Pan, Jae-Gu
; APPLICANT: Stoddard, Steven F.
; TITLE OF INVENTION: Sorbitol Dehydrogenases of Ketogulonigenium spp., Genes and Method
; FILE REFERENCE: 1533 3740000
; CURRENT APPLICATION NUMBER: US/10/162,713
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 2519
; TYPE: DNA
; ORGANISM: Ketogulonigenium sp.
US-10-162-713-9
Query Match 61.6%; Score 1071.2; DB 15; Length 2519;
Best Local Similarity 76.4%; Pred. No. 1.3e-308;
Matches 1329; Conservative 0; Mismatches 408; Indels 3; Gaps 1;
Qy 1 ATGAAACCCCAAGCGCTTTCGACAGCGGCGCGTGTCTATGCTTACCGGCGCCGCC 60
|||||

```



```

Db 750 ATGAATTCGATTGTTGCTTCTGGAAGGTTGCTGCGGTGATCTTTGCTGTCGCC 809
QY 61 GCATTGCGGAGTAACCCCGATTACCGATGAACCTGCTGGCAACCCCGCGCTGTGTA 120
Db 810 GCATTGCGGAGTAACCCCGATTACCGATGAACCTGCTGGCAACCCCGCGCTGTGTA 869
QY 121 TGAATTAACGCGGCAACAAGAAAATATGCGCACTGCGCCCTGACCCGATCACT 180
Db 870 TGAATTAACGCGGCAACAAGAAAATATGCGCACTGCGCCCTGACCCGATCACT 929
QY 181 GCCGACACGTTGTCATGTTGCACTGCTGGGCCCGCGGATGGAAGCGGCGGCTGA 240
Db 930 CCCGACACGTTGTCATGTTGCACTGCTGGGCCCGCGGATGGAAGCGGCGGCTGA 989
QY 241 CAGGTACGCGGATGATCCATGATGGGCTGATGATCTGGCAACCCCGGTGATGATC 300
Db 990 CAGGTACGCGGATGATCCATGATGGGCTGATGATCTGGCAACCCCGGTGATGATC 1049
QY 301 CAGGCGCTGGATGCGCAACAGGCGATCTGATCTGGAAACCCCGCGCACTGCGGCC 360
Db 1050 CAGGCGCTGGATGCGCAACAGGCGATCTGATCTGGAAACCCCGCGCACTGCGGCC 1109
QY 361 GTGCGCACTGAACCGCCCAAGGCGACCGGACCGCGGCTGCGCCCTTTACGCAAGAC 420
Db 1110 ACCTCGACGCTGAGCTGCTGGGGGATCGCAACGCGGCGATCGCGCTTTATGACCAAT 1169
QY 421 CTCTATTTCAGCTCATGAGCAACCACTGATCGCGCTGGATATGAGAAGGCGCAAGTC 480
Db 1170 GTCTACTCTGCTCTGTGGGCAACCACTGATCGCGCTGGATATGAGAAGGCGCAAGTC 1229
QY 481 GTATTGATGTCGACCTGATCGGGGAAAGACGCTTGAACCACTTACACACACGCGGCCG 540
Db 1230 GTCTTGCACGCTGACCGCGGCAAGGCGACGAGCGGCTGCG--AACTGCTCGCGCCC 1286
QY 541 ATTGCGCCATGCGGCTCATGCTGCGGGTTTCCACTGCGCATATTCGCGCTATGATGC 600
Db 1287 ATTGCGCCATGCGGCTCATGCTGCGGGTTTCCACTGCGCATATTCGCGCTATGATGC 1346
QY 601 TTATCTCGGGGCAAGATTCGCGGACGAGTGAAGAGCTGTGGGCAACCACTTATCCCG 660
Db 1347 TTATGTCGGGCAATGATGCAAGACGAGCGAAGAACTGTGGGCAACCACTTATCCCG 1406
QY 661 CAGCGCGGCAAGAGGCTGAAGACTTGGGGCAATGATTCAGAGCGGCTGATGATC 720
Db 1407 CAGCGAGGTAAAGGCTGAAGAACTGGGGCAATGATTCAGAGCGGCTGATGATC 1466
QY 721 GGGGCTGGGGTGAATCACTTATGATCCCGTGAAGCACTTGTGTTATGAGCTGAC 780
Db 1467 GGGGCTGGGGTGAATCACTTATGATCCCGTGAAGCACTTGTGTTATGAGCTGAC 1526
QY 781 GGGGTGGGCCCCAGCTTCCGAAACCCAGCGGCGACGCGGGCGGCACTGTATGACACC 840
Db 1527 GCGGTAGCGCCCGATCCGAGTTTCAACCGGCGACCCCGGCGGCACTGTATGACACC 1586
QY 841 AACACCGCTTTGCGGTGCTCCGACACCGGCGAGATTGTCTGAGGTCAACGACCTTG 900
Db 1587 AACACCGCTTTGCGGTGCTCCGACACCGGCGAGATTGTCTGAGGTCAACGACCTTG 1646
QY 901 CCGGCGACAACTGGGACCAAGATGCACTTGAAGTGTGCGCAACGTGATGTG 960
Db 1647 CCGGCGACAACTGGGACCAAGATGCACTTGAAGTGTGCGCAACGTGATGTG 1706
QY 961 CAAACCTCGCGGAGATGAGAGGCTGCGGCGCATCAACCCGATGCGGAGCGGCGAG 1020
Db 1707 CAAACCTCGCGGAGATGAGAGGCTGCGGCGCATCAACCCGATGCGGAGCGGCGAG 1766
QY 1021 CGCGGTGTGCTAGCGGCTGCGCTTTCAGAACCGGCGACGATGTGTGCTTTGATGCGGCC 1080
Db 1767 CGTGGCGTTGACCGGCGTTCCTGCAAAACCGGTTCATGTGTGGCGATTCGACGCTGAA 1826
QY 1081 TCGGGCGAATTCCTGTGGGCGGCTGATACCACTACCAATGATGATCGGCTGATGAC 1140
Db 1827 ACGGGCGAATTCCTGTGGGCGGCTGATACCACTACCAATGATGATGATGAC 1886

```

```

QY 1141 GAGACCGGCTTGTGACGCTGGAACGAGATGCGGTGCTGAAGAGCTGACGTTGATAT 1200
Db 1887 GAAACCGGCTTGTGACGCTGGAATGAAGATATCATCTTAAAGATCTGGACACCGACTAC 1946
QY 1201 GACGTCGCGCGACCTTCTCGGTGGTGGGCGGACCTGATGTCAGCGGACCTGAACCCGAC 1260
Db 1947 CGCATTTGCCGACATTTCTTGGTGGAGCGGACCTGCGGCGATCTTGAACCCCGAT 2006
QY 1261 ACCGCGATTTACTTCTTCCCGCTGAACATGCTGCTACGATATTTATGCGCGGTGATCA 1320
Db 2007 AGCGGACCTTACTTCTTCCCGCTGAACATGCTGCTACGATATTTATGCGCGGTGATCA 2066
QY 1321 GAGTTTACGCGCTGACGCTTATATACACGAGCGGACCGGCAAACTGCGCGCGGCTTT 1380
Db 2067 GAGTTTACGCGGATGAGGCTTATACACGAGCGGACCGGCAAACTGCGCGCGGCTTT 2126
QY 1381 GAAATATGCGCGCGGATGACGCGATTTGATATGACACCGGCGGACCTTGTGTCGCG 1440
Db 2127 GAAATATGCGCGCGGATGACGCGATTTGATATGACACCGGCGGACCTTGTGTCGCG 2186
QY 1441 GAGCGCGCTGCGGCAACTTCTGCGCGCTTTTGTGACGCGGCGGCTGATGTTCAAC 1500
Db 2187 GAGCGCTGCGGCAACTTCTGCGCGCTTTTGTGACGCGGCGGCTGATGTTCAAC 2246
QY 1501 GCGGGAACGACCGCTATTTCTGCGCTTCTGACGCAAGAACCGGCGAGACTTTGTGCGAG 1560
Db 2247 GCGGGAACGACCGCTATTTCTGCGCTTCTGACGCAAGAACCGGCGAGACTTTGTGCGAG 2306
QY 1561 GCGCGCTTGTGAGGCTGCGGACGCGGCGGCGGATGACGCTGATGATGAGCGGCTGCA 1620
Db 2307 ACCGCTGCGGACGCTGCGGACGCGGCTGATGACGCTGATGATGAGCGGCTGCA 2366
QY 1621 TATATGCGCATCGGTGCGGCGGCTGACCTATGAGCAAGCAATGAAACCGCGCTGCGC 1680
Db 2367 TATATGCGCATCGGTGCGGCGGCTGACCTATGAGCAAGCAATGAAACCGCGCTGCGC 2426
QY 1681 GAGGCAATGCAATGCACTCGGTGCTGATGCAATGATGCTTTGCACTGCGCGGCAATGA 1740
Db 2427 GCGGCAATGCAATGCACTCGGTGCTGATGCAATGATGCTTTGCACTGCGCGGCAATGA 2486

RESULT 5
US-10-162-713-3
; Sequence 3, Application US/10162713
; Publication No. US20030228672A1
; GENERAL INFORMATION:
; APPLICANT: Choi, Eun-Sung
; APPLICANT: D'Silva, John
; APPLICANT: Kim, Hye-Sun
; APPLICANT: Kim, Mi-Soo
; APPLICANT: Lee, Jung Kee
; APPLICANT: Pan, Jae-Gu
; APPLICANT: Stoddard, Steven F.
; TITLE OF INVENTION: Sorbitol Dehydrogenases of Ketogulonigenium spp.
; FILE REFERENCE: 1533 3740000
; CURRENT APPLICATION NUMBER: US/10162,713
; CURRENT FILING DATE: 2002-06-06
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1740
; TYPE: DNA
; ORGANISM: Ketogulonigenium sp.
US-10-162-713-3

Query Match 60.5%; Score 1052; DB 15; Length 1740;
Best Local Similarity 75.7%; Pred. No. 6,4e-303;
Matches 1317; Conservative 0; Mismatches 420; Indels 3; Gaps 1;
QY 1 ATGAACCCCAACAGCTGCTTGCACACGCGGCGGCTGATGCTTACCGCGCGGCC 60

```



QY 1 ATGACCCCAACAGCTCTTGGACACAGCGCGCGCTGTATGTCTTACCGCGCCGCG 60  
Db 1241 ATGAAAGCAGAGTCTTTTCTGTTTGAAGGCGTGTGCGCTTGGCAAGTACCGCAATTT 1300  
QY 61 GCATTGCGGAGGTAAACCCCGATTACCGATGAATGCTGCGCAACCCGCGCTGTGAA 120  
Db 1301 GCGCTTGCTATGTGAACCCCGTCAACGAGCTGCTGGCAAAACCGCGCGCGCGAA 1360  
QY 121 TGGATTACTAGGCGCGCAACCAAGAAACTATGCGCACTCGCCCTGACCCAGATCACT 180  
Db 1361 TGGATCAGTACCGCGCAACCAAGAAACTATGCGCACTCGCCCTGACCAAGATCAAG 1420  
QY 181 GCGGACAGCTGTGCTGCTGCAACTGTGTGGCGCGCGGAGTGAAGCGCGGCGCGTA 240  
Db 1421 CCGGAGAGCTCGGTGAGCTGCACTGTGTGGCGCGCGGAGTGAAGCGCGGCGAAAGTC 1480  
QY 241 CAGGTACCGCGATGATCATGATGCGGTGATGTATCTGGCAAAACCCGCGTATGTATC 300  
Db 1481 CAGGTACCTCGCTGATCCATGATGCGGTGATGTATCTGGCAAAACCCGCGCATCATC 1540  
QY 301 CAGGCGGTGATGCGCAACAGCGGATGTATCTGGGAAACCGCGCGCAACTGCGCGCG 360  
Db 1541 CAGGCGATCAGCTTAACCGGAGCTGATCTGGAAACACCGCGCGCACTGCGCGCAAC 1600  
QY 361 GTCGCGACGCTAAACGCGCAAGCGACCGCGACGCGCGCTGCGCCCTTAAACGCAACAGC 420  
Db 1601 GTGGCAACGCTGAACAGCTTCGATGAGCGGATCGCGGTATCGCGGTATGCGCAACAA 1660  
QY 421 CTCTATTACGCTCATAGGCAACCATCTATGCGCGTGTATGAGACGCGCGCAAGTTC 480  
Db 1661 GTTACTCTGCTGCGGGAACAACCTGTGTGCGTGTGACCAACCGCGCGCAAGTTC 1720  
QY 481 GTATTGATGCACTGTGATCGGCGGAGAGAGCGCTTGAACAGTAAACCAACGCGCGCG 540  
Db 1721 ACCTGACGCTGACCGCGCGCAAGGAGAGAGATGTTTC---TAAGTGTGCGCGCG 1777  
QY 541 ATTGTGCCAATGCGCTCATGCTGTGCGGCTTCACTGCGCAATATGCGCTATGATGCG 600  
Db 1778 ATGTGCTTAACGCGCGTGTGCGCGGCTTCACTGCGCAATATGCGCGCTGCGCTGCG 1837  
QY 601 TTATCTCGGCGGACGATCCGCGAGCGGTGAGAGAGTGTGCGCAACCTTATCCCG 660  
Db 1838 TTGCTTTGCGGCGATGACGCACTACCGGTGAAGAACTGTGCGCAACTTATCTCATCC 1897  
QY 661 CAGCGCGGCAAGAGGCTGACGAGCTTGGGCAATGATTTGAGGCGCGCTGATGAC 720  
Db 1898 AAGCGGCTGAAGAGGCGATGAACCTTGGGCAACGATACGAAACCGCGCTGATGAC 1957  
QY 721 GCGCTTGGGCTGATCACTATGATCCCGTGAACCACTTGTGTTTATGCTGAC 780  
Db 1958 GCGCTTGGGCGCAATCACTGACGACCTCGTACCAACTGTATTTTACGATGTG 2017  
QY 781 GCGGTGCGGCGAGCTCCGAAACCGCGCGGACGCGCGGCGCAACGCTGTATGCGAC 840  
Db 2018 GCGGTGCGGCGGCTTTCGAAACCGCAACGCGGCGCAACGCGGCGCAACTGATGCGCA 2077  
QY 841 AACACCGCTTGGGTGCGGTCCGACACGCGGAGAGTGTGTGCGGTCAACGACCTG 900  
Db 2078 AACACCGCTTGGGTGCGGTCCGACACCGGCGAAATCTGTGCGGTCAACCAATCTG 2137  
QY 901 CCGGCGGACACTGGGACCAAGATGACGTTGAGATGATGTGTGCGCAACGTCAGATG 960  
Db 2138 CCGGCGGACACTGGGACCAAGATGACGTTGAGATGATGTGTGCGCAATGTCAGCTG 2197  
QY 961 CAACCTTGGCGGAGATGAGGCTTGGCGCGCAATCAACCCCAATGCGCGAGCGCGAG 1020  
Db 2198 CAGCTTTCGCTGACATGACGCGCTGAATGATCAACCCCAACGCGCGCTGCGAG 2257  
QY 1021 CGCGGTGTGCTGAGCGGCTGCGCTTGAAGACGCGGCGAGTGTGTGATGCGGCG 1080  
Db 2258 CGTGTGCTGACCGCGCTTTCGCTGAACCGGTACCTATGTGCGAGTTCGACGCTGAA 2317  
QY 1081 TCGGCGAATTCCTGTGCGCGGCTGATACCAATCAACCAATATGATGCGCTGATGAC 1140

Db 2318 ACGGCGAATTCCTGTGCGCGCGGACGCACTACCAAAACATGATCAGTGCATGAC 2377  
QY 1141 GAGACCGGCTTGTGACGCTGACGAGATGCGGTGCTGAAAGCTGAGCGTTGAATAT 1200  
Db 2378 GAAACCGGCTGTGACGCTGATGAAGATATCACTCTAAAGATCTGACACCGACTAC 2437  
QY 1201 GACGCTGCCCGACCTTCCGTGGGTGGCGGACGCTGTGAGCGGACGCACTGAACCGGAC 1260  
Db 2438 GCGATTTCCCGCAATTTTGGGTGACGCACTGCGCGCTGCGCATCTTAAACCGAT 2437  
QY 1261 ACCGCAATTTACTTCTTCCGCTGAACATGCTGTAAGATATATATGCGCGGTGATCA 1320  
Db 2498 AGCGCATCTACTTCACTTCCCTGAACCAACGCTGTGCGGATTTGGCGGCGAGTCA 2557  
QY 1321 GAGTTAGCGGCTGACGCTATATACACGCGCGGACCGCAAACTCGCGCGGCTT 1380  
Db 2558 GAGTTCAAGGCAATGAGCGTGTACACACGCGCGACTTACTGCTTGGCGCGGAAAA 2617  
QY 1381 GAAATATGCGCGCGCATGACGCGATGATATACACCGCGCGCACTTGTGTGCGCG 1440  
Db 2618 GAAATATGCGCGCGCATGACGCGATGATATACACCGCGCGCACTTGTGTGCGCG 2677  
QY 1441 GAGCGCTTGGCGCACTACTCGCGCTTTTGTGACGCGCGAGCGCGGTGTTCAAC 1500  
Db 2678 GAGCGCTTGGCGCACTACTCGCGCTTCTGACGCGCGCGCGGTGCTGTCAAC 2737  
QY 1501 GCGCGGACGACCGGCTATTTCCGTGCGCGCGGAGAAACCGGAGACTTTGCGGAG 1560  
Db 2738 GCGCGGACGATCGCTTCTTCCGTGCGCGCGGAGAACTTGGCGAGCCCTGTGCGAG 2797  
QY 1561 GCGCGCTTGGCGACGCGTGCACGCGGCGAGCGATCACTGATGAGCGCGGTGCA 1620  
Db 2798 ACCGCTTGGCGACGCTGCTTCCGCGCGGAGCGCGGTGATGAGCGCGGTGAG 2857  
QY 1621 TATATCGCATCGGCTGCGCGGCGGTCTGACATGAGCAACGCAATTTGAACGCGCGCTGCGC 1680  
Db 2858 TACATGCGCATCGGCTGCGCGGCGGACCACTACGCGCGGCTGCAACCGTCCGTGCGC 2917  
QY 1681 GAGCGAATGATTCGACCTCGGTGCGGTATGCTATGCTTGTGCACTGCGCGAGTAA 1740  
Db 2918 GAGCGTGTGATGCTGACCTCGATGCGTAAACCGCGTGTACGTTTGTGCTGCGCGAGCA 2977

RESULT 7  
US-10-369-493-34527  
; Sequence 34527, Application US/10369493  
; Publication No. US20030233675A1

GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 34527  
; LENGTH: 2151  
; TYPE: DNA  
; ORGANISM: *Sphingomonas aromaticivorans*

Query Match 8.7%; Score 152.2; DB 15; Length 2151;  
Best Local Similarity 52.6%; Pred. No. 9,4e-35;  
Matches 477; Conservative 0; Mismatches 388; Indels 42; Gaps 5;

QY 83 TTACGATGAATGCTGTGCGGAAACCGCGCGCTGTGAATGATTAACGCGCGCAAC 142

```

Db      113 TCACCGACGCGCTCATGCGCCCAAGCGCCCAAGCGCAATGCGACTATGCGCGGATT 172
Qy      143 AAGAAAATCTATCGGCACTCGCCCTGACCGCAATCACTGCGCAACGTTGCTAGTGC 202
Db      173 ATGCGCGACGCGCTTCTCCCGCTCAACCGATCAATGCGCAACGTCGCGCACTCG 232
Qy      203 AACTGGCTGGGCGCGGATGGAAGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCG 259
Db      233 GTCTTGGCTGGTTCATGACCTGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 292
Qy      260 ATGATGCGCTGATGTATCTGCGAAACCGCGGATGATGATGCGCGCTGATGCGCAA 319
Db      293 ATGACGCGACGCTCTCATCTGACCGCGCTGCGATGATGAGCGGCTGATGCGCAAGA 352
Qy      320 CAGCGCATCTGATCTGCGAACAACCGCGCAATGCGCGCGCGCGCGCGCGCGCGCGCG 379
Db      353 CCGCGCGCTCAAGTGTCTTCAAGATCCGAAATCCGCGCGCGCGCGCGCGCGCGCGCT 412
Qy      380 AAGCGCA---CGCAACCGCGCGCTGCGCCCTTTACGCGCAAGCGCTCTATTTCACTCAT 436
Db      413 GCTGCGACGCGGTCAATCGCGCGCTGCGCGCTTATGCGCAAGGCTCTGCGCGACGC 472
Qy      437 GGGACAACTATGATTCGCGCTGATATGAGACGCGCGCGCGCGCGCGCGCGCGCGCG 496
Db      473 TCACGCGTGCCTGCTGCGCGCTGACCGAAGACCGAAAGGCTGCTGCTCCAAAGTGC 532
Qy      497 GTGATGCGCGCGAAGACGCGCTTACCAATACCAACCGCGCGCGCGCGCGCGCGCGCG 556
Db      533 TCGTGGCCAAACGAGGAGACTACAC---ATCACCGCGCGCGCGCGCGCGCGCGCGCG 589
Qy      557 TCAATGCTGCGGCTTCCACTGCGCAATTTCCGCTTATGATGATCTTATCTGCGGCG 616
Db      590 AGGTTCTGATGCGTACGCGCGCTCGAATCAAGCGCGCGCGCTTATGCGCGCTACG 649
Qy      617 ATTCCGCGACGCGTGAAGACTGTGCGCGCAACCACTTTATCCGCGCGCGCGCGAAGAG 676
Db      650 ACGTCAACACCGCGCAAGAGTGTGAAGTTCACACCGCTCCCGCGCAATCCAGCGGAG 709
Qy      677 GTACGAGACTTGGGCGAATGATTTGAGGCGCGCTGATGACGCGC-----723
Db      710 GATTTCGAGAACAGGCGATGAGAAACGCGCGCGCGCGCGCGCTGCGCGAATGATGAGAG 769
Qy      724 -----GTTGGGCTCAATCACTATATCCCGGAGAACTTTGAT 766
Db      770 TCGCGCGGCGCGCGCGTGTGAGATTCATCACTATATCCCGCGCAACCTTCTGC 829
Qy      767 TCTATGCTGACCGCGGTGCGCCAGCGTCCGAAACCGACGCGCGCGCGCGCGCGCG 826
Db      830 TGTTCGCGACCGCGCATGCGGACCATGGAACCGCGCGCGCGCGCGCGCGCGCG 889
Qy      827 CGCTGATGCGCAACACCGCGTTTGGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCG 886
Db      890 GCGCTCTA---CACTGCTGATGCTGCGGTGATCCGATACCGCGCGCTATGCTGCG 946
Qy      887 GTACACGAGCCCTGCGCGCGCGCAACTGCGGACCAAGATGCACTTTCAGATGATGCTG 946
Db      947 ATTTCAGAGAAACCGCGAAGACGCGTGGGACTTTCAGCTCCGCGCGAGATCAGCTG 1006
Qy      947 CCAACGT 953
Db      1007 CCGACCT 1013

```

## RESULT 8

```

US-10-369-493-45426
; Sequence 45426, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.

```

```

; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 45426
; LENGTH: 1731
; TYPE: DNA
; ORGANISM: Rhodobacter capsulatus
US-10-369-493-45426

```

```

Query Match      5.7%; Score 99.4; DB 15; Length 1731;
Best Local Similarity 44.5%; Pred. No. 4,8e-19;
Matches 739; Conservative 0; Mismatches 881; Indels 41; Gaps 7;

```

```

Qy      104 ACCGCGCGCTGTGAATGATTAATACGCGCGCAACCAAGAAACTATGCGCACTCGC 163
Db      80 AGCAGACCGCGCGCGAGTGTGACGACGAGATGCGCGCGCGCGCGCGCGCGCGCG 139
Qy      164 CCGTACCCAGATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 216
Db      140 CGCTTGCACGCTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 139
Qy      217 --CCGCGATGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 274
Db      200 TGGCGCGGAGAAACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 259
Qy      275 ATTCGCGAAACCGCGGATGATGATCAAGCGCGCTGATGCGCGCGCGCGCGCGCGCG 334
Db      260 ACATACCGGCTCTATTCGCGGCTTTACCGCGATGACATCAACCGCGCGCGCGCG 319
Qy      335 GCGAACAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 394
Db      320 GCGAATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 376
Qy      395 GCGCGTGCCTTTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 454
Db      377 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 436
Qy      455 CGCTGATGAGAGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 514
Db      437 GCGTGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 496
Qy      515 GCTTACCAATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 574
Db      497 GCTATTC---TACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 553
Qy      575 CCGCGCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 634
Db      554 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 613
Qy      635 AGCTGTGCGCGCAACCACTTTATCCCGCGCG-----CGCGCGAAGAGGTGA 680
Db      614 TGTCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 673
Qy      681 CGAAGCTTGGGCGAATTTTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 740
Db      674 CGATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 732
Qy      741 CTATGATCCGCTGACGAACTTGTGTTATGCTGACCGCGCGCGCGCGCGCGCGCG 800
Db      733 GCGACCTGGGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 792
Qy      801 AACCGACGCG-----GACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 852
Db      793 AACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 852
Qy      853 GCGGTGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 912

```

```

Db      853 GGGCACAACCTTAAAGCCGCTCGCCGCTGGGAGTCAACCCGGCAGCGGAGATCAAG 912
Qy      913 TGGGACCAAGATATGACGTTTCAGATATGTCGCCAATCTGCATATGCAACCTCGGCC 972
Db      913 TGGGACCTTCAAGACACCCCGCGAGGGCTGGGATTTTCAACGGCGTGAACGAGTGGT 972
Qy      973 GAATATGAGGATCTGCGCGCATCAACCCCAATGCGGCGAGCGGCGAGCGCTGTGCTG 1032
Db      973 GGGCTTGTTCACAAGACGCGCAACAGCGCTTTGCCACCCCGCAACGCAACGCTTTCTTT 1032
Qy      1033 ACGGATGCGCTTTCACAAGACGCGCAACGATGTGTCGTTTATGTCGGGCTCGGCGAATTC 1092
Db      1033 TATGTGTCAACCGCGGAGCGCAATTTGTCAACGCCCTCGCTTTCGGAAGAACATC 1092
Qy      1093 CTGTGGGCGCGTATACCACTACACCAATATATATGCTCGATCAACGAGACGAGCGCTT 1152
Db      1093 ACCGTGGGCGCAAGGACATCGATGAACCGGCGCTCCGATCTACAGAGAGAACCGCCCC 1152
Qy      1153 GTACGAGTGAACGAGATGCGGTGCTGAAGAGCTGGACGTGAATATAGCTTCGCTCCCG 1212
Db      1153 GGTGCCCCCATATGCGCGGCGCAAGCGCCAGAGGCGCAAGATCTTTGCGCTGCCCCG 1212
Qy      1213 ACCTTCCTGGATGCGCGCACTGTGCTCAAGCGCACTGAACCCGCAACCGGCAATTTAC 1272
Db      1213 ---TTCTGCGGCGCAAGACTGATGCGATGCGCTACAGCGCAACGAGGCGCTCTTC 1269
Qy      1273 TTCTTGCGCTGAACCAATGCTGCTAGCATATTTATGCGCTGTATCAAGAGTTTACGCG 1332
Db      1270 TATGTCTCCCTCGAACGATGGGAGTGTATGTGAAACGAGCCGATCAATTAACAAGAG 1329
Qy      1333 CTGACGCTATTAACACCAAGCGGCGCAAAACTCGCGCGGCGCTTTGAAATATAGGCG 1392
Db      1330 GGGCGGCGCTTACCTGGGCGGCGGCTTACGATCAAGCGCTCTTCAAGAGTTTACATCGCG 1389
Qy      1393 CGCATGACCGCATTTGATATCAAGACCGGCGCACTTGTGTGCGCGAGCGCGCTGCG 1452
Db      1390 TCGCTCAAGCGCATCGATTCGAATACGGGCGCGGTGAATGAGATCAAGAACGAGCGG 1449
Qy      1453 GCGAACACTACGCGCGCTTTGTGACGCGGCGGCGGTGTGTTCAACGCGGAGACGAGC 1512
Db      1450 CCGCTGTGGGCGGCGGTGATGACACCGCGGCGCAAGCTGTCTTCAACGCGCAACCCGAG 1509
Qy      1513 CGCTATTTCCGTCGCTCCTCAACGCAAGAACCGCGGAGACTTTGTGCGAGCGCGCTTGGG 1572
Db      1510 GGGGAGTTTCAAGCGCTGACGCGGCAACCGCGGAAAGCTGTGTTCGACGACGCGC 1569
Qy      1573 ACGGTCCGCAACGGGCGAGCGGATCACTACGATTAAGATTTGAACGCGCTGCAATTTATCGCATC 1632
Db      1570 TCGGAGCATGCTCGACCAAGCGATCACTGGAAACAGACGCGGCAATATGTCTCGATC 1629
Qy      1633 GGTGCGGCGGCTGACCTATGACGCAATTGAACGCGCGCTGGCCGAGCAATCAT 1692
Db      1630 ATCTCGGCGGTGGGCGGCGCGGTGCGGCTTTGGGCGGGAAGTGGCAAGGCGCTCAAC 1689
Qy      1693 TCACACTCGGTGGTATGATCTATGTCTTTCGACTGCC 1733
Db      1690 TATCTGAACCAAGGCGGACGAGCTGTGACCTTCGCGCTGCC 1730

```

## RESULT 9

```

US-10-369-493-28294
; Sequence 28294, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493

```

```

; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 28294
; LENGTH: 1677
; TYPE: DNA
; ORGANISM: Burkholderia fungorum
US-10-369-493-28294

Query Match      4.8%; Score 83.8; DB 15; Length 1677;
Best Local Similarity 47.1%; Pred. No. 2.3e-14;
Matches 438; Conservative 0; Mismatches 467; Indels 24; Gaps 5;

Qy      38 TGTATTTGCTTAAACGCGCGCGCGGCTATTCGCGAGATTAACCCGATTAACGATGAATGC 97
Db      32 TGTATTTGATTCGGGCGCGGCGCGGCGGCGGATTAACCGGCTACCTATGAGGTTTGA 91
Qy      98 TGGCAACCGCGCGCTGTGATATGATTAATTAACGGCGGCAACCAAGAAATATATGCC 157
Db      92 CAGCTGCGAAGACGACCCCGCTGTGCTACGTACTACCGACGTACACGCGCAAGCGCGC 151
Qy      158 ACTGCGCCCTGACCCAGATCACTGCGCAACAGCTTGTGCTAGTTGCAACTGCTGCGGCC 217
Db      152 ATTGCGCGCTCAAGAGATTCACCGCGAGCGTGAAGAACCTCAAGAGGTCTGAGCT 211
Qy      218 GCGGATGAAGCGG-----GCGCTACAGTCAAGCGCGATGATTCATATGCGC 268
Db      212 ACAATTTTCGCGGATCTGACAGAGGCTTTCGAAGCGCACCCGATGTCATCAACGAGCT 271
Qy      269 TGATGTATCTGACAAACCCCGGTGATGATCAAGCGCGCTGATGCGCAACAGCGCATC 328
Db      272 ACCTTTTGATACACGCGCGAAAGAACAGTCTATGATTTGATGCGCGCACCGCAAGC 331
Qy      329 TGATTTGGAACACCGCGCGCACTGCGCGCTGCGCAAGCTTAACGCCAAGCGCAAC 388
Db      332 AGTTGTGAAGATTCGAACGAGCTCGGTCGCGAATCGTTCAACGCGCATGTTGCGAGC 391
Qy      389 G---CAAGCGGCGGTGCGCCCTTTACGCGACAGACCTCTATTTGACATGAGGACAAAC 445
Db      392 TGATTAACCGGCGGTGCGCTGTACCGCAAGAAATGCTACGTGCGAGTGTGAGCGCG 451
Qy      446 ATCTGATCGCGCTGATATGAGACGCGCGCAGCTGATTTGATGATGCAAGTGAATCGG 505
Db      452 ACGTGTGCGCTCGACGCGCGACGAGCGCGCGCTGCGCTGCGAAGAGATGTTGCGAC 511
Qy      506 GCGAAGACGCTTGAACAGTAACACACAGGCGCGATTTGATGATGCAAGTGAATCGG 565
Db      512 CCGGCTTGAGTA---GCGCTTCGCTGCGCGCGCTGCGCTGACGCGCGCACTGATGG 568
Qy      566 CCGGTTCCACTGCGCAATATCGCCCTATGATGATCTTTATCGGCGCAACATTCGCGCA 625
Db      569 TGGTAGGCGAGGCGGCAATACGCGCGCGCGCGCTTTTATGCGCGCATGTAATCGGATA 628
Qy      626 CCGGTGAAGACTGTGCGCAACATTTATTCGCGACGCGCGGCGAAGAGGTGACGAGA 685
Db      629 ACGGAATGTCTATGAAAGCGCTTCAAGGTGCGCGCGCGCGGGAAGGCGCGCAACA 688
Qy      686 CTGCGGCAATGATTTGAGGCGCGCTGATGACGCGGCTGCGGCTCGATGACCTATG 745
Db      689 GGTGCGCAACGCGATGAGAAACAGCGGCGCGCGCGCTGCTGACGCGCACTTACG 748
Qy      746 ATCCGTCAGCAACTTGTGTCTATATGCTGACGCGCGGTGAGGCGACGCTCGCAACCC 805
Db      749 ACGGAGCATGAAACCTTTACTGGGCGCGGCAACCCCGGCGCGCTGCTCGCGATC 808
Qy      806 AGCGGCGAGCGCGGCGGCGCGCTGATGAGCAACAACCGCTTGGCGGTGCTGCGCG 865
Db      809 TGCG-----ACGCGGACACACTTAT---TTCGATTGCTACTGCACTCGATCGCA 859
Qy      866 ACAAGGCGAATTTGCTGCGGTGACCAAGACCTTCCGCGGACACTGAGACCAAGAT 925
Db      860 AAACGCGGATCTCAATGAGCATACAGATGACGAAGAACATGACACATGGGACTACGAG 919

```



QY 402 CGCCCTTACGCGACGAGCCTCTATTTCAGCTCATGGAGAACCATCTGATCGCGTGA 461  
DB 420 CGCGCTGTACAAAGGCGAAAGTGTTCGCGCGCTTATGACGCGCGCTGTGCGCTCA 479  
QY 462 TATGAGACGAGGCGCAGGCTGATTCATGATGTCGAAACGTGATTCGGGAGAAACGCTTGA 521  
DB 480 TGCCGCGACCGGAGCAAGCGCTGGGAGATGACACGCTGATGACACACGACCTCTCA 539  
QY 522 CAGTAACACCAACGAGGCGGATTCGCAATGCGATGATGCGCGGCTTCCACTGCA 581  
DB 540 CACCATACCGGTCGCGCGCGGCTGTCACGCGCAAGTGTGATGGGAAACGCGCGCG 599  
QY 582 ATATTCGCTTATGATGCTTATCTTCGAGGACGATTCGCGACGCGGTGAGAGCTGTG 641  
DB 600 CAAATACGCGCGCGCTGATGCTGACGCTTATGATGCGGACAGGCAAGCAGGCAATG 659  
QY 642 GC-----GCAACCACTTTATCCGCGAGCGGCGAAGA 674  
DB 660 GCGCTGGTTACGGTTCGCGGCGCATCCGAGCAAGCGCTTGAAGACGATCGATGAGA 719  
QY 675 GGGTACAGACCTTGGGCGCAATGATTCAGAGCGCGCTGATGACCG-----CGT 725  
DB 720 GCGCGCGCAACCTGGGATCCGCGCGCAATGATGCTCAATGCGGTCGCGCGCACCGC 779  
QY 726 CTGGGGGTCAATGATCCTGATCCCGTGAAGCAACTTGTGCTATGCGTTCGACCGCGCT 785  
DB 780 GTGGAGACCATGATCTTCGATCCGACCTCACTGATCTGATGTCGCGCACGCGCAACG 839  
QY 786 GGGCGCGACGCTCCGAAACCGAGCGCGACCGCGCGCGGCGGCACTGTATGACCAACAC 845  
DB 840 CTGCGCGTGAATCGGATCTGCGCAGACCGCGCGGTCGCAACATCTGATCTGCGCTC 899  
QY 846 CGCGTTTGGCGTCCGTCGCGACGAGCGCGGATTCCTGGCGTTCACCAACCTCGCGCG 905  
DB 900 GATGTCGCGCTGATGCTGCGACCGCGCAATGCTGCTGCGACTATCAGAGAGCGCGCG 959  
QY 906 CGACAACTGGGACCAAGATGCACTGATGATGTCGCGCAAGTGCATGTCGACAC 965  
DB 960 CGACAACTGGGATTAATCTGCGACCGCGGATGATCTCGGGAATGCGCATGCGACGG 1019  
QY 966 CTGCGCGGAGATGAGGCTTCGCGCGC 992  
DB 1020 CAAGCGCGCGCAAGTATCTGCGACGC 1046

## RESULT 12

US-10-369-493-31374  
; Sequence 31374, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 31374  
; LENGTH: 1596  
; TYPE: DNA  
; ORGANISM: Burkholderia cepacia  
US-10-369-493-31374

Query Match 4.3%; Score 74.8; DB 15; Length 1596;  
Best Local Similarity 48.7%; Pred. No. 1e-11;  
Matches 264; Conservative 0; Mismatches 272; Indels 6; Gaps 2;

QY 248 CGCCGATGATCCATGATGCGGTGATGATCTGGCAAAACCCCGGTATGATCCAGGCGC 307  
DB 182 CGCGCATGATTCGAACGCGGTGATGTCATTAACCACTCGTTAATCACTGATGAGG 241  
QY 308 TGGATGCGCAAAAGCGCATCTGATCTTGGGAAACACCGCGCACTGCGCGCGCTGCA 367  
DB 242 TCGATGCGGTGACGCGCAAGAAATTCGCACTACAGCAACAAGATGGCGCGCTCACTA 301  
QY 368 CGCTAAAGCCCAAGGCGCACGCGGCGCTGCGCTTTCAGGCGACGACCTCTATT 427  
DB 302 CGTTTGTTCG---GCGCGCAACATCGCGCGGTGCGCATATGCGCGGCGACCGTCTCA 358  
QY 428 TCACTCATGGAACAACCATCTGATTCGCTGGAATGAGACGCGCGCATGCTATTG 487  
DB 359 TGGGACCTTCGACGCGCAAGCTTCGCGCTGCGACGCAAAAACCGGACGCTGTTGCGC 418  
QY 488 ATGTGCAACGTGATTCGGGCGAAGCGGCTTGAACAGTAACACCAACGCGGCGCATTCG 547  
DB 419 AAGCGCATGTCGCGATCCCATGAAGCTATTGCG---AAGCATGCGCGCGACCTG 475  
QY 548 CCAATGCGGTATCGTCGCGGCTTCACCTGCGCAATATTCGCGCTATGATGCTTTACT 607  
DB 476 TCGACGCAAGTGTGATTCGCGCACGAACGCGCGCAATATGATGATTCGCGTTCTCA 535  
QY 608 CGGCGCAAGATTCGCGCAACGCGGTGAGAGCTGTGCGCAACCATTTATCCGCGACCGG 667  
DB 536 AGGCTTTCGACGCAACCTCCGCTCAATTCGTGGAACCTTTATACGATTCGGAACCG 595  
QY 668 GCGAAGAGGTGACGAGACTTGGGCAATGATTTTCAGGCGCGCTGATGACCGCGCTCT 727  
DB 596 GCGAGGAAGTGTGTGGGCGCAAGAGACGCTACCGCGCGGAGGAGCGCAACATCG 655  
QY 728 GGGTCAATGATCTATATCCCGTGAAGAACTTGTGTTATGAGCTTCACCGCGCTG 787  
DB 656 ACGCGAAGAAAGACAGCTTCGCGAAGAGCGCGCACTTCTACAGAAGCTTCGCGCGC 715  
QY 788 GC 789  
DB 716 GC 717

## RESULT 13

US-10-369-493-28616  
; Sequence 28616, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 28616  
; LENGTH: 2051  
; TYPE: DNA  
; ORGANISM: Burkholderia fungorum  
US-10-369-493-28616

Query Match 4.3%; Score 74.8; DB 15; Length 2051;  
Best Local Similarity 48.7%; Pred. No. 1.1e-11;  
Matches 264; Conservative 0; Mismatches 272; Indels 6; Gaps 2;



QY 308 TGGATGCGAAACAGGCGATCTGATCTGGGAAACACCGCGCACTGCGCGCGTCCGCA 367  
DB 338 TCATCGCGTGAAGCGGCAAGATTTCTGGCACTCAAGACAAAGATGGCGCGGTCACTA 397  
QY 368 CCGTAAACCGCCCAAGCGACCGCAAGCGCGTCCGCTTTACCGGACGACCTCTATT 427  
DB 398 CGTTTGTTCG--GGGCCCAACATCGCGCGTGGCCATAGCCCGGCAACGTTCTTACA 454  
QY 428 TCAGCTCANTGGGCAACCACTGATCGCGTGGATATGAGACGGGCCAGTCTGATTTCG 487  
DB 455 TGGGCAACCTTCGACCGGAAAGCTGTGCGCTCGACGCAAAAACCGGACCGTGTGCG 514  
QY 488 ATGTCAACGTGTGATCGGGGGAAGACGCGTTGACCACTAACCAACGCGGCGATTTGCG 547  
DB 515 AAACGAGATCGCGCATCCCGATGAAGCTATTTCG--AAACGATGGCGCGGACCGTGG 571  
QY 548 CCATGCGGCTCATCTGTCGCGGTTTCACCTCGCAATTTGCGGCTATGATCTTATCT 607  
DB 572 TCACGCGCAAGGTGTTGATCGGACGAGCGCGGCAATATGCGATTCCGCGTTCTCA 631  
QY 608 CCGGCGACGATTCGCGGACGCGGTGAGAGCTGTGCGCAACCACTTATCCGCGACCGG 667  
DB 632 AGCGGTTTCAGCGAAATCTCGGTCAATGTGTGACCTTCTATACGATTCGGAACCG 691  
QY 668 GCGAAGAGGTGACGAGACTTGGGGCAATGATTTGAGGCGCGCTGGATGACCGGCTCT 727  
DB 692 GCCAGGAGGTGTGTGGGCGACAGAGACGCTACCGCGCGGACGGGAAAGCGGACATCG 751  
QY 728 GGGGTCGATGACCTATGATCCCGTACGAACTTTGTGTTATGAGCTCGACCGGCGTGG 787  
DB 752 ACCGCCGAGAAAAGACGCTTCCGAGAAAGCGCGCACTTCTACAGAGCTCGGCGGCG 811  
QY 788 GC 789  
DB 812 GC 813

RESULT 14  
US-10-369-493-31733  
; Sequence 31733, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; SEQ ID NO 31733  
; LENGTH: 1869  
; TYPE: DNA  
; ORGANISM: Rhodobacter sphaeroides  
US-10-369-493-31733

Query Match 4.3%; Score 74; DB 15; Length 1869;  
Best Local Similarity 49.6%; Pred. No. 1.8e-11;  
Matches 249; Conservative 0; Mismatches 245; Indels 8; Gaps 2;

QY 1201 GAGCTGCGCGGACCTTCGCGGAGCGGAGCTGTCGACCGGCACTGAACCGGAGC 1260  
DB 1246 GGGCTTCGCGCGGCGCGGCTCGGCTCCAGAGCAAGACCGGCTCTCTTCGCGCAAG 1305  
QY 1261 ACCGGCATTTACTTCTTCCGCGTGAACATGCTG-----CTAGCATATTAAGCGCTT 1314  
DB 1306 ACCGGCTCTCTATATGTCGAGCAACCACTCTGATGAGCTACAGAGCCCTACCGGCTC 1365

QY 1315 GATCAGAGTTTACGCGGCTCGACGTTCTATTAACACGCGCGAACCGCAAACTCGCGCG 1374  
DB 1366 AGTACTCGCGCGGCGGCGGCTTATGTCGCGGCCACCTCTGATGTATCGCGCGCAAC 1425  
QY 1375 GCGTTTAAATATATGCGCGGCGCATCGACGGAATTGATATGACGCGGCGGCACTTGTGG 1434  
DB 1426 AGCCATGCGCGCATGCGCAATTTATGCTCGGGAACCGACGAAAGGCGGACATCAATGG 1485  
QY 1435 TCGCGGAGGCGCGCTCGCGGAACTACTCGCCGTTTGTGACGCGGACGCGGTGTGTTG 1494  
DB 1486 TCCTGCGCGGAGAGCTTCTCGGTGTGTTGCGGCGCTCTGCGCACCGCGGCTGACGTGTG 1545  
QY 1495 TTCAACGCGCGGACCGCAACCGCTATTTCGCGGCTCTGACCGGAAACCGGCGGACCTTGG 1554  
DB 1546 TTTCACGCAACGCTCGAAGCTATCTGAAGCGCGTCAATCGGAGACGGCGAGCTGCTC 1605  
QY 1555 TGGCAGCGCGGCTCTTTCGACGCTCGACGCGGCGGCGGCGATACGCTACGAGTTGACGCG 1614  
DB 1606 TATCGCTTCAAGACCGCGCTCGCGGCTCATCGGCAAGTATGACTTACGATTCACAGGCG 1665  
QY 1615 GTGCAATATATCGCATCGATCGGTCGCGGCGGTCTGACCTATGCGACGCAATGAAACGCGG 1674  
DB 1666 AAGCATATGTGGCGGCTGTCTCGGCGG--TCGCGGCTGGGCGGCGCATGCGCTCGCG 1723  
QY 1675 CTGGCGGAGGCAATCATGATCGA 1696  
DB 1724 CCGGCTGACCAACCCCAACGA 1745

RESULT 15  
US-10-369-493-34525  
; Sequence 34525, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; SEQ ID NO 34525  
; LENGTH: 2120  
; TYPE: DNA  
; ORGANISM: Sphingomonas aromaticivorans  
US-10-369-493-34525

Query Match 4.0%; Score 68.8; DB 15; Length 2120;  
Best Local Similarity 45.9%; Pred. No. 6.7e-10;  
Matches 431; Conservative 0; Mismatches 467; Indels 42; Gaps 4;

QY 95 TGCTGGGAAACCCCGCGCTGTGATGATTAATCAAGCGCGGCAACGAAATATATC 154  
DB 100 TCTTGACCGAGGCGCGGAGCGAGTGGCGCGCGCGGCTGAGCGCGGCAAGACGC 159  
QY 155 GCGACTGCGCGCTTACCACTCACTGCTCCGACGAGCTTGTCAAGTTGCACTGCTGGG 214  
DB 160 ATCACTCGCGCTTACCGCATCAACGCGGAGACGTCGCTGCGCTTGGCTGGCGTGC 219  
QY 215 CCGCGGATGAGAGCGGCGCGCTAC--AGTCAAGCGGATGATCATGATGCGTGA 271  
DB 220 AAGTGAACCTTGGACCTTGCGCGGAGAGGAGGAGACCGCGGTGTGTGTGAGAGCGTGC 279  
QY 272 TGTATCGGCAACCCCGGTATGATTCAGGCGCTGAGTTCGCAACGCGGATGTA 331  
DB 280 TCTACACTCGGAGCAACGCGCGCGCTATGTGCTTGTGATGCGGCGACCGGCAAGGAGC 339



QY 332 TCTGGAAACACCCGCGCCAACTGCGCGCGCTGCGCACGCTTAAAGCCCAAGCGACG-- 389  
Db 340 TGTGGCGGTTGAGCCCGAGGTGACATGACAGTCAACCGCACGCTCTGCTGCACATG 399  
QY 390 -CAAGCGCGGCGTGCAGCCCTTTACGGACAGACGCTTATTTCAAGCTCATGGGACACCATC 448  
Db 400 TCAACCTGGCGTGCAGCGGTGGGGCGCGCAAGCTCTTCTGCGCACTTCGATGGGTGA 459  
QY 449 TGAATCGGCTGATATGAGACGCGCCAGGTGCTATTCGATGCAAGTGAATCGAGC 508  
Db 460 TGTACGCACTGATGCGCGACCGCGGCTGCTGAAAGCGACTTCATCAGAGACC 519  
QY 509 AAGACGCTTGAACCAATCAACGAGGCGGATGTCGCAATGGCGTATCTGCGG 568  
Db 520 GAGCCAAAGGCGCAATTCCACCGCGCTCCGAAATCGGCGGCGATGTCGTCATCG 579  
QY 569 GTTCACCTGCAATATTCGCGCCATGATGCTTATCTCGGGACAGATTCGCGACG 628  
Db 580 GCATGGCGGCGCGGAGTACGAGCTGCGCGCTATGTCAGCGGCTGATCTGATACCG 639  
QY 629 GTGAGAGCTGTGGCGCAACCACTTATCCC-----GCAGC 664  
Db 640 GAAAGCTGCGCTGCGCTTGGCAGTGTGCCCCGCGATCCCAAGCTGAGGCGCGAGAAA 699  
QY 665 CGGCGGAAAGGAGTGAAGACTTGGGGCAATGATTCGAGGCGCGCTGATGACCGGCG 724  
Db 700 CGCCGAACTGAAAGCGCGCTGAAGACTTGGAGCCGAAACAGCCGCTGGGACATGCGG 759  
QY 725 TC-----TGGGGTCAAGTCACTATGATCCGTGACGAACTTGTGTTCTATG 772  
Db 760 GCGGTGCGAGCCGTTGGATGCGATCAACTAGACCCGAAACCGGCTGTCTTGTG 819  
QY 773 GCTGAGCGGCGTGGGCGCCAGCGTCCGAACCCAGCGCGGACGCGGCGGACGCTGT 832  
Db 820 GCAACGGCAATGGCGGCTTTACGCCACGTCCAAGCGCTGCTGCGGGGGCGACACC 879  
QY 833 ATGACCAACACCCGCTTTCGAGTGCATCCGACACGAGCGAGATTGTCGCGTCAAC 892  
Db 880 TCTACCTCGCCAGGCTGTGCGGTGATCCGAAGACCGGCGGATGAAGTGCACATAAC 939  
QY 893 AGACCTGCGCGCGAGAACTTGGAGCCAGAAATGCAAGTTGAGATGATGTTGCCAAG 952  
Db 940 AGGAAAGCGCGGCGCAACTTGGAGCTTCAACCGCAACCGCGGATGATCTCACCCGCA 999  
QY 953 TCGATGTCAAACCTCGCGCGAGATGAGGAGTGGCGCGC 992  
Db 1000 TGAAGATCGACGCGGAGGACCGGCCAAGTCTGCTCCACGC 1039

Search completed: November 22, 2004, 20:44:06  
Job time : 943 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 2004, 14:15:00 ; Search time 5784 Seconds  
(without alignments)

10962.160 Million cell updates/sec

Title: US-10-802-682-4

Perfect score: 1740  
1 atgaaccacacacgcgtctgct.....tcttgcactgcgcagctaa 1740

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hic:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gss1:\*  
9: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	58.8	3.4	941	BZ564301	BZ564301 pacas2-164
C 2	57.8	3.3	751	BZ561554	BZ561554 pacas2-164
C 3	57.2	3.3	580	AZ935093	AZ935093 BJ_BA000
C 4	54.6	3.1	1014	BZ561555	BZ561555 pacas2-164
C 5	51.8	3.0	517	CA298989	CA298989 SCMCV103
C 6	51.8	3.0	742	CA272117	CA272117 SCVPLB208
C 7	51.6	3.0	858	BZ56371	BZ56371 pacas2-164
C 8	51.4	2.9	925	CNS0091P	AL053013 Drosophila
C 9	50.2	2.9	658	CA133332	CA133332 SCORL103
C 10	50.2	2.9	701	CA190595	CA190595 SCORL103
C 11	48.6	2.8	647	CA150641	CA150641 SCBFR204
C 12	46.2	2.7	791	BH378255	BH378255 AG-ND-142
C 13	46.2	2.7	815	BH400384	BH400384 AG-ND-160
C 14	46	2.6	639	CA107253	CA107253 SCRFH100
C 15	46	2.6	709	CL659599	CL659599 PRI0134C
C 16	45.6	2.6	298	CAZ577200	AZ577200 04903 Sho
C 17	45.6	2.6	723	BZ549537	BZ549537 pacas1-60
C 18	45.4	2.6	925	CNS0091P	AL053013 Drosophila
C 19	45.2	2.6	659	CA131657	CA131657 SCBGR104
C 20	45.2	2.6	1191	BZ555577	BZ555577 pacas1-60
C 21	45	2.6	640	CA270833	CA270833 SCVPLB208
C 22	44.4	2.6	783	CC729915	CC729915 OGUCO11TH
C 23	44.4	2.6	845	CC659386	CC659386 OGUXH61TH
C 24	44.4	2.6	856	CG325844	CG325844 OGW3074TH

25	44.4	2.6	861	9	CC659397	CC659397 OGUXH61TV
C 26	44.4	2.6	862	9	CG325853	CG325853 OGW3074TH
27	43.8	2.5	650	6	CA244227	CA244227 SCBFL408
28	43.6	2.5	559	4	BG739469	BG739469 EM1_81_CO
29	43.6	2.5	578	2	BE592327	BE592327 MS1_93_C1
30	43.6	2.5	613	2	AW678546	AW678546 MS1_16_F0
31	43.6	2.5	625	4	BG739296	BG739296 EM1_81_CO
32	43.6	2.5	681	2	BE593953	BE593953 MS1_104_D
33	43.4	2.5	763	6	CA212935	CA212935 SCBCL7C0
34	43.4	2.5	935	9	CNS006XK	AL066051 Drosophila
35	43.2	2.5	690	6	CD448001	CD448001 EL01N0203
36	43.2	2.5	696	6	CD866097	CD866097 AZ02_102J
37	43.2	2.5	728	9	CC672860	CC672860 OGJBP74TV
C 38	43.2	2.5	795	9	CC262434	CC262434 OGJBP74TV
C 39	43.2	2.5	821	9	CC672836	CC672836 OGJBP74TH
C 40	43.2	2.5	889	9	CNS004NB	AL054280 Drosophila
C 41	43.2	2.5	839	9	CG334632	CG334632 OGOC47TV
C 42	43.2	2.5	932	9	CNS0072Q	AL066742 Drosophila
C 43	43.2	2.5	1007	9	CC673577	CC673577 OGJUM38TV
C 44	43	2.5	718	5	BH294429	BH294429 603605349
C 45	43	2.5	793	9	CG234022	CG234022 OGZAP76TH

## ALIGNMENTS

RESULT 1  
BZ564301/c 941 bp DNA linear GSS 17-DEC-2002  
LOCUS pacas2-164\_4680.y2 pacas2-164 Pseudomonas aeruginosa genomic clone  
DEFINITION pacas2-164\_4680, genomic survey sequence.  
ACCESSION BZ564301.1 GI:27189813  
VERSION BZ564301.1  
KEYWORDS GSS.  
SOURCE Pseudomonas aeruginosa  
ORGANISM Pseudomonas aeruginosa  
REFERENCE Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,T.L., Kaul,R. and Oiser,M.V. Whole-genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library J. Bacteriol. (2002) In press  
JOURNAL Contact: Chris K. Raymond  
COMMENT Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: cgraymond@u.washington.edu  
Class: shotgun.  
FEATURES  
source location/Qualifiers  
1. 941  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="2-164"  
/db\_xref="taxon:287"  
/clone="pacas2-164\_4680"  
/clone\_lib="pacas2-164"  
/note="Clinical isolate 2-164 Whole genomic shotgun library."  
ORIGIN  
Query Match 3.4%; Score 58.8; DB 8; Length 941;  
Best Local Similarity 46.4%; Pred. No. 0.00039;  
Matches 192; Conservative 0; Mismatches 222; Indels 0; Gaps 0;  
QY 1202 ACCTGTGCGCCGACCTTCTGTGCGCGCCGACCTGTGCGCGCCGACCTGACCAACCCGACCA 1261  
DB 490 AAGTATCACCCCGTCTCTGCGCGCGCAAGAACTGAAACCCGATGCGCTACGCGCAAGACA 431  
QY 1262 CCGGATTACTCTTGTGCGCGCTGAACAATGCTGCTACGATATATATGCGCGCTGATCAAG 1321

```

Db      430 CCGGCTGTTCTTACGTCGCGCCCAACCACTGAGAGAAACTACTGAGACCGAGAGGTCA 371
Qy      1322 AGTTTACCGGCTGTCGATCTTATACACAGCGCCACCCGAAACTCGCGCGCTTGG 1381
Db      370 GCTATACGAAAGGCGACCGCTTACCTTGGCATGGGCTTCCGATCAACGCGATGTAGAGC 311
Qy      1382 AAAATATGGGCGCGCATCGACGCGATGATATGACGACCGGCGCACTTGTGTGGCGCG 1441
Db      310 ACCAGCTCGGACCTGCGCGCATGAGACCGGTCAGCGGCAAGTGTGTGGAAACACA 251
Qy      1442 AGCGCCCTGGGCGAACTACTGCGCGCTTGTTCGACGCGAGCGGTGTGTGTTCACG 1501
Db      250 AGGAACACCTGCGCTGCGGCGCGGCTGCTGCGCCACCGCGCACTGTGTTCACCG 191
Qy      1502 GCGGAGCCGACCGCTATTTCCGTGCGCTCAGCCAGAAACCGGCGGACATTTGTGCGAG 1561
Db      190 GCAACCGGCGAGCGCTACTTCAAGGCTTTCGACCGCAAGAGCGGCAAGAGCTGTGAAAT 131
Qy      1562 CCCGCTTTCGACGCTGCGGACGCGGCGACGCGATGACGATGAGATTGACGCGCG 1615
Db      130 TCCAGACCGGCGACGCGCATGCTGCTCGCACCGATCACTGGGAAAGAGCGGCG 77

```

```

RESULT 2
LOCUS   BZ561554 751 bp DNA linear GSS 17-DEC-2002
DEFINITION pac2-164_3321.x1 pac2-164 Pseudomonas aeruginosa genomic clone
VERSION BZ561554
KEYWORDS BZ561554.1 GI:27181766
SOURCE  GSS.
ORGANISM Pseudomonas aeruginosa
          Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
          Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 751)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
          Burns,J.L., Kaul,R. and Olsen,M.V.
          Whole-Genome-Sequence variation among multiple isolates of
          Pseudomonas aeruginosa library
          J. Bacteriol. (2002) In press
          Contact: Chris K. Raymond
          Genome Center
          University of Washington
          Box 352145, Seattle, WA 98105-2145, USA
          Tel: 2062216954
          Fax: 2066857244
          Email: craymond@u.washington.edu
          Class: shotgun.
FEATURES
          source
            1..751
              location/Qualifiers
                /organism="Pseudomonas aeruginosa"
                /mol_type="genomic DNA"
                /strain="2-164"
                /db_xref="taxon:287"
                /clone="pac2-164_3321"
                /clone_1lb="pac2-164"
                /note="clinal isolate 2-164 whole genomic shotgun
                library."
ORIGIN
  Query Match 3.3%; Score 57.8; DB 8; Length 751;
  Best Local Similarity 49.9%; Pred. No. 0.0008;
  Matches 208; Conservative 0; Mismatches 197; Indels 12; Gaps 2;

```

```

Db      161 GCGCCCTGACCGCATCTGCGGACGATGTTGTGATGTCAGTGTGTGTGCGC----- 215
Qy      81 CCGCGCTCAAGCAAGTCACGCGGACGACGCTGTTCAGAGCTGACCCCGGCTGTGTACT 140
Db      216 ----CGCGGAGATGAGAGCGCGGCGCTACAGGTCAAGCGCGATGATCATGATGCGTGA 271
Qy      141 CTTTCGCGACGAGAAACAGCGCGCGGACAGAGTCCAGGCGATCTGACGAGCGCGTGA 200

```

```

Qy      272 TGTATCTGCAAAACCCGCGATGTGATCCAGCGCTGATGCGCAACAGCGCATCTGA 331
Db      201 TCTACGTGACCGCATCTTACTTCGCGGCTGTTCGCCCTCGAGAGGAGACCGGAGCGCC 260
Qy      332 TCTGGGAACACCGCGGCAACTGTCGCGCGGTGCGCACGCTAAACGCCCAAGGACCGCA 391
Db      261 TGTGACCTTCAACACACCGCGCTGCGGACGACATCCGCGCGTGTGC---GACGTGATCA 317
Qy      392 AGCGGCGCTGCGCCCTTATACGACGACGCTCTATTTCAGCTGATGAGGACCAACATCTGA 451
Db      318 ACCGCGCGCGCGCATTTATGCGGACAGGTGTCTTTCGCGACCCCGATCGGTCG 377
Qy      452 TCGCGCTGATATGAGACGCGGCGAGTGTGATTCGATGTCGACGCTGATCGGCGAG 511
Db      378 TGGCGCTGAACAGAACACCGCGAGGTGTGTGAGAGAGAGTTTCGCGATCAGCGCG 437
Qy      512 ACCGCTTGAACAGTAAACACGCGGCGGATGTGCGCAATGCGCTCATGTCGCGG 568
Db      438 CCGGCTACACCATGATCTGCGCGCGCGCACCTGTGTAAGAGCGGCAAGACCGGAGG 494

```

```

RESULT 3
LOCUS   A2935093/c 580 bp DNA linear GSS 24-APR-2001
DEFINITION BT_Ba0003G14r B. japonicum BAC library Bradyrhizobium japonicum
          genomic, genomic survey sequence.
VERSION A2935093
KEYWORDS A2935093.1 GI:13777242
SOURCE  GSS.
ORGANISM Bradyrhizobium japonicum
          Bradyrhizobium japonicum
          Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
          Bradyrhizobiaceae; Bradyrhizobium.
REFERENCE 1 (bases 1 to 580)
AUTHORS Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A.,
          Goicoechea,J.L., Stacey,G., Sadowsky,M.J. and King,R.A.
          A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
          Genome
          Genome Res. 11 (8), 1434-1440 (2001)
          21376150
          11483585
          Contact: Wing RA
          Clemson University Genomics Institute
          Clemson University
          100 Jordan Hall, Clemson, SC 29634, USA
          Tel: 864 656 7268
          Fax: 864 656 4293
          Email: twing@clemson.edu
          Class: BAC ends
          High quality sequence stop: 541.
FEATURES
          source
            1..580
              location/Qualifiers
                /organism="Bradyrhizobium japonicum"
                /mol_type="genomic DNA"
                /strain="USDAL110"
                /db_xref="taxon:375"
                /lab_host="E. coli"
                /clone_1lb="B. japonicum BAC library"
                /note="Vector: pindigo536; Site_1: HindIII"
ORIGIN
  Query Match 3.3%; Score 57.2; DB 8; Length 580;
  Best Local Similarity 50.8%; Pred. No. 0.00095;
  Matches 193; Conservative 0; Mismatches 178; Indels 9; Gaps 2;

```

```

Qy      118 GAATGATTTACTGCGGCGCGAACCAAGAAATATGCGCATCGCCCTGACCCAGATC 177
Db      383 GACTGGCTGACTTATTCACGCTTCTTATAGTCTTATATCACTACAGTCCCTTGGCATC 324
Qy      178 ACTCGCAACAAGTGTGTCAGTTGCAACTGTGTGCGCC-----CGCGGATGAGGCG 231
Db      323 AATACGAGCAATGTGGTAACTTAAGCGGTGCGGTGATTCATATTCGCGGAGCGTCCACC 264

```

```

QY 232 GGGGCGGTACAGGTACAGCCGATATCCATGATGCGCTGATGATGTTTGGCAAAACCCCGGT 231
Db 263 CGTGGCTTGCAAGTCAAGCCGCTTGCGGCGGACGCGGTGCTTACTACAGAGGCTCTTAC 204
QY 292 GATGTGATTCAGGCGCTGATGATGCGCAACAGAGCATGATCTGTGGAACACCGCGCGCA 351
Db 203 AGTGAAGTGTGGGCTGAACGCGGACAGGAGGTGATCTGTGTTCTTATTTCCCGAA 144
QY 352 CTGCCC--GCGTGCACGCTAAACGCCCAAGCCACCGCAAGCGCGCGTGCCTT 408
Db 143 CTTGACGAAGCCCTGATGCGCGGACAGCGACTCTCCCTATATATCGCGCGCTGCGCTC 84
QY 409 TAGGCAACGAGCCCTTATTTAGCTCATGAGGACAAACCATGATGCGGCTGATATGAG 468
Db 83 GGAAGAGGCAAGTCTTACGTGGCAACATGATGCGCGCTCATGCGCTTGATATGAG 24
QY 469 ACGGCGCAGGTCTGATTGCA 488
Db 23 ACTGAAAGTCCGCTGGA 4

RESULT 4
B2561555/c 1014 bp DNA linear GSS 17-DEC-2002
LOCUS pacs2-164_3321.y2 pacs2-164 Pseudomonas aeruginosa genomic clone
DEFINITION pacs2-164_3321, genomic survey sequence.
ACCESSION B2561555
VERSION B2561555.1 GI:27181769
KEYWORDS GSS.
ORGANISM Pseudomonas aeruginosa
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1014)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
source Location/Qualifiers
1..1014
/mol_type="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_3321"
/clone_id="pacs2-164"
/note="clinical isolate 2-164 whole genomic shotgun
library."
ORIGIN
Query Match 3.1%; Score 54.6; DB 8; Length 1014;
Best Local Similarity 48.8%; Pred. No. 0.0048;
Matches 147; Conservative 0; Mismatches 154; Indels 0; Gaps 0;
QY 1381 GAAATATGGCCGCGCATGACGCGATTGATATGACACCGGCGCACTTGTGTCGCG 1440
Db 395 GACACGTCGCGACCTGCGGCGCATGACCGCGTACGCGCAAGTGTCTGGAAAC 336
QY 1441 GAGCGCCCTCGGCGGAATCTAGCCCGTTTGTTCAGCGGACGCGGTGTGTTTCAAC 1500
Db 335 AAGGAACACCTGCGGCTGCGGCGGAGTCTGCGCACCGCGCAACCTCGTGTTACCC 276
QY 1501 GCGCGGACGACGCTATTTCGTCGCTCCTCAGCCAGGAACCGGCGAGACTTGTGGCAG 1560

```

```

Db 275 GGACCGGAGACGGCTACTTCAAGGCTTTCGACGCGGAAGCGGAGAGCTTGGA 216
QY 1561 GCCCGTCTTCCAGCAGGTGCGGACGCGGACGCGATCAGCTAGAGTTGGACGCGCTGCA 1620
Db 215 TTCAGACCGGAGGAGGAGTCTGTCGACCGGATCACTGGGAACAGACGCGACG 156
QY 1621 TATATGCCATGATGTCGCGGCGGTCTGACCTATGCAAGCAATTGAACGCGCGCTGCGC 1680
Db 155 TACCTCGAGTAGAAGCTGCGTACGGAACGCGCGTGTGTGGGCGGCGACATGGCC 96
QY 1681 G 1681
Db 95 G 95

RESULT 5
CA299889 517 bp mRNA linear EST 26-SEP-2003
LOCUS SCMCIV1031F09.g LV1 Saccharum officinarum cDNA clone SCMCIV1031F09
DEFINITION 5', mRNA sequence.
ACCESSION CA299889
VERSION CA299889.1 GI:36073038
KEYWORDS EST.
SOURCE Saccharum officinarum
ORGANISM Saccharum officinarum
Saccharum officinarum
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum
complex.
REFERENCE 1 (bases 1 to 517)
AUTHORS Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
TITLE The libraries that made SUCEST
JOURNAL Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT Centro de Biologia Molecular e Engenhariaia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: paruda@unicamp.br
Contact: Arruda P
Clone distribution: clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
http://www.bcccenter.fcav.unesp.br
Plate: 031 row: F column: 09
Seq primer: T7 Promoter Primer.
FEATURES
source Location/Qualifiers
1..517
/mol_type="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCMCIV1031F09"
/lab_host="DH10B"
/clone_id="LV1"
/note="Organ: Etiolated leaves from in vitro grown
seedlings; Vector: pSPori1; Site_1: SalI; Site_2: NotI; An
unidirectional cDNA library generated from [Etiolated
leaves from in vitro grown seedlings]. cDNA was prepared
from polyA+ mRNA using SuperScript Plasmid System Kit
(Invitrogen). The double-strand cDNAs were fractionated
in a sepharose CL-2B 40cm-columns and fragments sizing
between 0.8 and 1.5 Kb were directionally cloned into the
vector. Details of each source of RNA and library
construction can be obtained at
http://succest.iad.ic.unicamp.br/public"
ORIGIN
Query Match 3.0%; Score 51.8; DB 6; Length 517;
Best Local Similarity 53.1%; Pred. No. 0.023;
Matches 110; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
QY 814 ACGCGGCGGACGCTATGACCAACACCGCTTGTGCGGTGCTCCGACACGCGC 873

```

Db 261 ACAGACGACAGGGGCGGCGCTGACAGGACGCGCGGTGACTACAGCTTTCGCGGCC 320  
 QY 874 GAAATTGTCTGGGTGACACGACACCTCGCCGCGGAACTGGGACCAAGATGACAGTTTC 933  
 Db 321 GACCTGGGACATGACCGCGAGAGCCCGGACGCGGACACCTCGGAGCTTCCCAAGGTGCC 380  
 QY 934 GAGATGATGTCCTCCCAACGTGATGTGCAACCTTCGCGCGAGATGAGGGATCGCGGCC 993  
 Db 381 GCGCTGACGGGCGCTCAACAACGCGGACACCGCGTGTACGACATCAAGGCGCTGCGGCC 440  
 QY 994 ATCAACCCCAATGCGGCGGACGCGCGAG 1020  
 Db 441 ATGAGCGCTGTGTGCGCGCGCGCG 467

RESULT 6  
 CA272117 742 bp mRNA linear EST 26-SEP-2003  
 LOCUS CA272117  
 DEFINITION SCVPLB2086D05.g LB2 Saccharum officinarum cDNA clone SCVPLB2086D05  
 5', mRNA sequence.

ACCESSION CA272117 GI:35983014  
 VERSION CA272117  
 KEYWORDS EST.  
 SOURCE Saccharum officinarum  
 ORGANISM Saccharum officinarum

REFERENCE  
 AUTHORS Sakayota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum  
 complex.

1 (bases 1 to 742)  
 Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.  
 The libraries that made SUCEST  
 Genet. Mol. Biol. 24 (1-4), 1-7 (2001)  
 COMMENT Contact: Arruda P  
 Centro de Biologia Molecular e Engenharia Genetica  
 Universidade Estadual de Campinas  
 Caixa Postal 6010, 13083-970, Campinas SP, Brazil  
 Tel: 55 19 3788 1137  
 Fax: 55 19 3788 1089

EMAIL: parudu@unicamp.br  
 Clone distribution: clone distribution information can be found  
 through the Brazilian Clone Collection Center (BCCC) at  
 http://www.bcccenter.fcav.unesp.br  
 plate: 086 row: D column: 05  
 Seq primer: T7 Promoter Primer.  
 Location/Qualifiers

1. 742  
 Location/Qualifiers

/organism="Saccharum officinarum"  
 /mol\_type="rRNA"  
 /db\_xref="taxon:4547"  
 /clone="SCVPLB2086D05"  
 /lab\_host="DH10B"  
 /clone\_id="LB2"  
 /note="Organ: Lateral buds from plants adult plants  
 growing in greenhouse; Vector: pSport1; Site: 1; Salt;  
 Site 2: NotI; An unidirectional cDNA library generated  
 from [lateral buds from plants adult plants growing in  
 greenhouse]. cDNA was prepared from polyA+ mRNA using  
 SuperScript plasmid System Kit (Invitrogen). The  
 double-strand cDNAs were fractionated in a sepharose  
 CL-2B 40cm-columns and fragments sizing between 0.8 and  
 1.5 Kb were directionally cloned into the vector. Details  
 of each source of RNA and library construction can be  
 obtained at http://sucestr.lad.ic.unicamp.br/public"

## ORIGIN

Query Match 3.0%; Score 51.8; DB 6; Length 742;  
 Best Local Similarity 52.0%; Pred. No. 0.025;  
 Matches 116; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 814 ACAGCGGCGGACGCTGTATGACACACCGCTTTCGCGGTGCGTCCCGACAGGCG 873  
 Db 147 ACAGCGTGCAGGGCGGCGGACGACGCGCGCTGTGACTACGCGCTTTCGCGGCC 206

QY 874 GAGATGTCGCGCTACACGACCTCGCGCGACACACTGGAGACCAAGATGACAGTTTC 933  
 Db 207 GAGCTGGGACATGACACCGCGAGAGCCCGGACGCGGACACCTCGGAGCTTCCCAAGGTGCC 286  
 QY 934 GAGATGATGTCCTCCCAACGTGATGTGCAACCTTCGCGCGAGATGAGGGATCGCGGCC 993  
 Db 267 GCGCTGACGGGCGCTCAACAACGCGGACACCGCGTGTACGACATCAAGGCGCTGCGGCC 326  
 QY 994 ATCAACCCCAATGCGGCGGACGCGCGAGCGCGGTGTGCTGACGG 1036  
 Db 327 ATGAGCGCTGTGTGCGCGCGCGCG 369

RESULT 7  
 B2566371 858 bp DNA linear GSS 17-DEC-2002  
 LOCUS B2566371  
 DEFINITION \*pac62-164\_6041.y2 pac62-164 Pseudomonas aeruginosa genomic clone  
 pac62-164\_6041, genomic survey sequence.

ACCESSION B2566371 GI:27195788  
 VERSION B2566371  
 KEYWORDS GSS.  
 SOURCE Pseudomonas aeruginosa  
 ORGANISM Pseudomonas aeruginosa

REFERENCE  
 AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
 Burns,J.L., Kaul,R. and Olsen,M.V.  
 Whole-Genome-Sequence variation among multiple isolates of  
 Pseudomonas aeruginosa library  
 J. Bacteriol. (2002) In press  
 COMMENT Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: ckraymon@u.washington.edu  
 Class: shotgun.

FEATURES  
 source Location/Qualifiers

1. 858  
 Location/Qualifiers  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="2-164"  
 /db\_xref="taxon:287"  
 /clone="pac62-164\_6041"  
 /clone\_id="pac62-164"  
 /note="Clinical isolate 2-164 Whole genomic shotgun  
 library."

## ORIGIN

Query Match 3.0%; Score 51.6; DB 8; Length 858;  
 Best Local Similarity 49.6%; Pred. No. 0.028;  
 Matches 132; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 1468 GTTGTGACGAGCGAGCGGTGTGTCTCAACGCGGAGACCGACTATTTCCGTGCC 1527  
 Db 565 GTTGTGCGCACACCGCTCAACCTGTGTCAACGCGGACGAGCGGCTACTTCAAGGCC 506  
 QY 1528 CTCAGCCAGGAACCGGAGACTTTGTGTGAGGCCCGGTCTTGTGAGCGGTGAGCGGG 1587  
 Db 505 TTGACGCGGAGAGCGGAGAGAGACTGTGAAATTCAGACGCGGACGCGGATCGTTCG 446  
 QY 1588 CAGCGATCAGCTACGATTGAGACGCGGCTGCAATATATCCCATCGGTGCGGGGCTTG 1647  
 Db 445 CCACCGATCACTCTGGGAACAGAGCGGAGAGTACTCGGCGTGAACCGCTACGCGC 386  
 QY 1648 ACCTATGACGCAATTTGAAGCGCCCGCTGACCGAGCAATTCATTGACCTCGTGGT 1707  
 Db 385 GCGCGCGGCGCGCTGTGGGCGGAGACATGAGCGGCTGACCGCGCGGTGGCCAGAGGC 326  
 QY 1708 AATGATCTATGCTTTGACACTGCC 1733

Db 325 GATTCTTGGGATTCAAGCTGCC 300

RESULT 8  
CNS0091P/c  
LOCUS  
DEFINITION  
CNS0091P 925 bp DNA linear GSS 03-JUN-1999  
Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACR19D16 of RPL1-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
GSS.  
AL053013.1 GI:4934461  
Drosophila melanogaster (fruit fly)  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 925)  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
Aaron Mammoeer in Pictet de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPL1-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
P1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
source  
1..925  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACR19D16"  
/clone\_lib="RPL1-98"  
/note="end : TET3"

ORIGIN  
Query Match 3.0%; Score 51.4; DB 9; Length 925;  
Best Local Similarity 14.8%; Pred. No. 0.032;  
Matches 56; Conservative 163; Mismatches 160; Indels 0; Gaps 0;

210 CTGGGCGCGGGGATGAGGCGGGGCGGTACAGTCCGATGATCCATGATGCGCT 269  
Db 924 SBBSCSCSSBSCSSBSSSTSSBSCSSBSSSTSSBSSSTSSBSSSSSSSS 865  
Qy 270 GATGTATCTGTGCAAAACCCCGGTGATGATCAGGCGCTGATCCGCAACAGGCGATCT 329  
Db 864 SGTSSACVKNMASSSCGCGCMAABCCMCSSSSSCGSAABGVKXRAAGAGKRGSG 805  
Qy 330 GATGTGGAACACCGCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 389  
Db 804 GAGASHSSSSSACSSSSSSCSASCMASSSSSSSSSSSSSSSSGGAGGSGASSSSSSA 745  
Qy 390 CAAGCGGGCGGTGCGCTTAAAGCAGCAGAGCTCTATTACGTCTAGGAGCAACCATCT 449  
Db 744 SAGSVSSASSSCCT 685  
Qy 450 GATGCGCTGATGATGAGACGGCGCAGGTGATTCGATGATGAACTGATGCGGCA 509  
Db 684 SWSGCTSSASMSAARSS 625

Qy 510 AGACGCTTACCACTAACACACCGGCGGCGATTTCGCCATGCGCTCATCGTCCGG 569  
Db 624 MSSGGSSSVSASMSMSSSSSSGGSSGSGGAGGSSGSSGSSGSSGSSGSSGSS 565  
Qy 570 TTCACCTGCCCAATATTCG 588  
Db 564 GCMCRCSGSSAAATAACV 546

RESULT 9  
CA133332  
LOCUS  
DEFINITION  
SCQERT1033G08.g RT1 Saccharum officinarum cDNA clone SCQERT1033G08  
5' mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
EST.  
CA133332.1 GI:35019216  
Saccharum officinarum  
Saccharum officinarum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD  
clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum  
complex.  
1 (bases 1 to 658)  
Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.  
The libraries that made SUCEST  
Genet. Mol. Biol. 24 (1-4), 1-7 (2001)  
Centro de Biologia Molecular e Engenhariaia Genetica  
Universidade Estadual de Campinas  
Caixa Postal 6010, 13083-970, Campinas SP, Brazil  
Tel: 55 19 3788 1137  
Fax: 55 19 3788 1089  
Email: parnada@unicamp.br  
Clone distribution: clone distribution information can be found  
through the Brazilian Clone Collection Center (BCCC) at  
<http://www.bcccenter.fcav.unesp.br>  
Plate: 033 row: G column: 08  
Seq primer: T7 Promoter Primer.

FEATURES  
source  
1..658  
/organism="Saccharum officinarum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4547"  
/clone="SCQERT1033G08"  
/lab\_host="DH10B"  
/clone\_lib="Rn1"  
/note="Organ: Root tips (0.3cm-long) from adult plants;  
Vector: pSPori1; Site\_1: SalI; Site\_2: NotI; An  
unidirectional cDNA library generated from (Root tips  
(0.3cm-long) from adult plants). cDNA was prepared from  
polyA+ mRNA using Superscript Plasmid System Kit  
(Invitrogen). The double-strand cDNAs were fractionated  
in a sepharose CL-2B 40cm-columns and fragments sizing  
between 0.8 and 1.5 Kb were directionally cloned into the  
vector. Details of each source of RNA and library  
construction can be obtained at  
<http://sucest.fad.ic.unicamp.br/public>

ORIGIN  
Query Match 2.9%; Score 50.2; DB 6; Length 658;  
Best Local Similarity 51.6%; Pred. No. 0.063;  
Matches 115; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

814 AGCGCGGCGGCGGCTGTATGAGCACAACACCGCTTGGCGGTGCGCGACGAGCG 873  
Db 205 ACGAGTGTCAAGGGGGGCGGAGACAGGCGGCGGTGAGCTAGGCTTGTGCTGGCC 264  
Qy 874 GAGATTGCTGCGGTACACCAACCTCTGCGCGCGGCACTGGAGCAAGATGACGTT 933  
Db 265 GAGCTGGGAGATGACCGGAGAGGCGCGGACGACCTGGGGCTCTGCCAAGTGGCC 324  
Qy 934 GAGATGATGTGCGCAACGTGATGTGCAACCTCGCGCGGAGATGAGAGGTCTCGCGCC 993

Db 325 GCGCTACGGGCGCTCAACAACGCCGACACCGCCGTGTACACATCAAGGCCCTGCTGCC 384  
 QY 994 ATCAACCCCAATGCGGCGAGCGCGCGCCGTGTGTGTAACGG 1036  
 Db 385 ATGAGCGCTGGCGCGGATGGCCCGACGCGCGCGCGCTGG 427

RESULT 10  
 CA190595 701 bp mRNA linear EST 24-SEP-2003  
 CA190595  
 LOCUS 5', mRNA sequence.  
 DEFINITION SCCCRTIC03609.g R1 Saccharum officinarum cDNA clone SCCCRTIC03609

ACCESSION CA190595  
 VERSION CA190595.1 GI:35134909  
 KEYWORDS EST.  
 SOURCE Saccharum officinarum  
 ORGANISM Saccharum officinarum

## REFERENCE

1 (bases 1 to 701)  
 Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.  
 The libraries that made SUCEST  
 Genet. Mol. Biol. 24 (1-4), 1-7 (2001)  
 CONTACT: Arruda P  
 Centro de Biologia Molecular e Engenharia Genetica  
 Universidade Estadual de Campinas  
 Caixa Postal 6010, 13083-970, Campinas SP, Brazil  
 Tel: 55 19 3788 1137  
 Fax: 55 19 3788 1089

## COMMENT

Email: parnuda@unicamp.br  
 Clone distribution: clone distribution information can be found  
 through the Brazilian Clone Collection Center (BCCC) at  
<http://www.bcccenter.fcav.unesp.br>  
 Plate: C03 row: G column: 09  
 Seq primer: T7 Promoter Primer.  
 Location/Qualifiers

## FEATURES

source

1..701  
 /organism="Saccharum officinarum"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4547"  
 /clone="SCCCTIC03609"  
 /lab\_host="DH10B"  
 /clone\_idb="R21"  
 /note="Organ: Root tips (0.3cm-long) from adult plants;  
 Vector: pSport1; Site\_1: SalI; Site\_2: NotI; An  
 unidirectional cDNA library generated from [root tips  
 (0.3cm-long) from adult plants]. cDNA was prepared from  
 polyA+ mRNA using Superscript Plasmid System Kit  
 (Invitrogen). The double-strand cDNAs were fractionated  
 in a sepharose CL-2B 40cm-columns and fragments sizing  
 between 0.8 and 1.5 Kb were directionally cloned into the  
 vector. Details of each source of RNA and library  
 construction can be obtained at  
<http://sucest.lad.ic.unicamp.br/public>

## ORIGIN

Query Match 2.8%; Score 50.2; DB 6; Length 701;  
 Best Local Similarity 51.6%; Pred. No. 0.063;  
 Matches 115; Conservative 0; Mismatches 108; Indels 0; Gaps 0;  
 QY 814 ACCGCGGCGGCAACGCTGTATGCGACCAACCGCTTTCGGTGTCCGACAGGAGC 873  
 Db 205 ACAGCGTGCAGGGGGGGGGGAGAGGAGCGCGGTGACTACGCGTTTCTGGCGGCC 264  
 QY 874 GAGATTGTCTGGCGTCAACGACCGCTGCGCGGCAACTGGGACCAAGATGCAAGTTTC 933  
 Db 265 GAGCTGGGCAATGACCGCGAGAGCGCGGACGACCTGGGCGCTTCCGCAAGTGGCC 324  
 QY 934 GAGATGATGTGGCAACGCTGATGTGCAACCTCGCGCGAGATGAGGAGGTCTGCGCGCC 993

Db 325 GCGCTACGGGCGCTCAACAACGCCGACACCGCCGTGTACACATCAAGGCCCTGCTGCC 384  
 QY 994 ATCAACCCCAATGCGGCGAGCGCGCGCCGTGTGTGTAACGG 1036  
 Db 385 ATGAGCGCTGGCGCGGATGGCCCGACGCGCGCGCGCTGG 427

RESULT 11  
 CA150641 647 bp mRNA linear EST 24-SEP-2003  
 CA150641  
 LOCUS SCBFR22045E02.g R22 Saccharum officinarum cDNA clone SCBFR22045E02  
 DEFINITION 5', mRNA sequence.  
 ACCESSION CA150641 GI:35053896  
 VERSION CA150641.1 GI:35053896  
 KEYWORDS EST.  
 SOURCE Saccharum officinarum  
 ORGANISM Saccharum officinarum

REFERENCE 1 (bases 1 to 647)  
 Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.  
 The libraries that made SUCEST  
 Genet. Mol. Biol. 24 (1-4), 1-7 (2001)  
 CONTACT: Arruda P  
 Centro de Biologia Molecular e Engenharia Genetica  
 Universidade Estadual de Campinas  
 Caixa Postal 6010, 13083-970, Campinas SP, Brazil  
 Tel: 55 19 3788 1137  
 Fax: 55 19 3788 1089

## COMMENT

Email: parnuda@unicamp.br  
 Clone distribution: clone distribution information can be found  
 through the Brazilian Clone Collection Center (BCCC) at  
<http://www.bcccenter.fcav.unesp.br>  
 Plate: 045 row: E column: 02  
 Seq primer: T7 Promoter Primer.  
 Location/Qualifiers

## FEATURES

source

1..647  
 /organism="Saccharum officinarum"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4547"  
 /clone="SCBFR22045E02"  
 /lab\_host="DH10B"  
 /clone\_idb="R22"  
 /note="Organ: Shoot-root transition zone from young plants  
 (small insert library); Vector: pSport1; Site\_1: SalI;  
 Site\_2: NotI; An unidirectional cDNA library generated  
 from [shoot-root transition zone from young plants (small  
 insert library)]. cDNA was prepared from polyA+ mRNA  
 using Superscript Plasmid System Kit (Invitrogen). The  
 double-strand cDNAs were fractionated in a sepharose  
 CL-2B 40cm-columns and fragments sizing between 0.8 and  
 1.5 Kb were directionally cloned into the vector. Details  
 of each source of RNA and library construction can be  
 obtained at <http://sucest.lad.ic.unicamp.br/public>

## ORIGIN

Query Match 2.8%; Score 48.6; DB 6; Length 647;  
 Best Local Similarity 51.1%; Pred. No. 0.16;  
 Matches 114; Conservative 0; Mismatches 109; Indels 0; Gaps 0;  
 QY 814 ACCGCGGCGGCAACGCTGTATGCGACCAACCGCTTTCGGTGTCCGACAGGAGC 873  
 Db 131 ACAGCGGCGAGGGGGGGGAGAGGAGCGCGGTGACTACGCGTTTCTGGCGGCC 190  
 QY 874 GAGATTGTCTGGCGTCAACGACCGCTGCGCGGCAACTGGGACCAAGATGCAAGTTTC 933  
 Db 191 GAGCTGGGCAATGACCGCGAGAGCGCGGACGACCTGGGCGCTTCCGCAAGTGGCC 250  
 QY 934 GAGATGATGTGGCAACGCTGATGTGCAACCTCGCGCGAGATGAGGAGGTCTGCGCGCC 993  
 Db 251 GCGCTGACGGGCGTCAACAACGCGCAACGCGGTGTATGACATCAAGGCGCTGCTGCC 310



QY 994 ATCAACCCCATGCGGACGCGGACGCCCGCTGTGCTACGG 1036  
 Db 311 ATGAGCGCTGATGCGCGATGCGCCGACGCGCGCGCTGG 353

## RESULT 12

LOCUS BH378255/c 791 bp DNA linear GSS 10-DEC-2001  
 DEFINITION AG-ND-142L4.TR ND-TAM Anopheles gambiae genomic clone AG-ND-142L4,  
 genomic survey sequence.

ACCESSION BH378255  
 VERSION BH378255.1 GI:17324397  
 KEYWORDS GSS.  
 SOURCE Anopheles gambiae (African malaria mosquito)  
 ORGANISM Anopheles gambiae

REFERENCE 1 (bases 1 to 791)  
 AUTHORS Hong Y.S., Hogan J.R., Wang X., Sarkar A., Sim C., Loftus B.J.,  
 Ren C., Huff B.R., Carlile J.L., Black K., Zhang H.-B.,  
 Gardner M.J. and Collins F.H.

TITLE Construction of a BAC library and generation of BAC end  
 sequence-tagged connectors for genome sequencing of the African  
 malaria mosquito Anopheles gambiae  
 JOURNAL Mol. Genet. Genomics 268 (6), 720-728 (2003)

MEDLINE 22542063  
 PUBMED 12655398

COMMENT Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: bjoftus@igr.org

FEATURES  
 source  
 1..791  
 /organism="Anopheles gambiae"  
 /mol\_type="genomic DNA"  
 /strain="PEST"  
 /db\_xref="taxon:7165"  
 /clone.lib="AG-ND-142L4"  
 /clone.lib="ND-TAM"  
 /note="Vector: pCBAC1; Site\_1: HindIII"

## ORIGIN

Query Match 2.7%; Score 46.2; DB 8; Length 791;  
 Best Local Similarity 64.5%; Pred. No. 0.7;  
 Matches 69; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 110 CCGCTGTGAATGATTAATTAATGAGCCGCAACCAAGAAACTATGCCACTGCCCTCGA 169  
 Db 194 CCCCCGCGATTTGAGAACTAGGAGCAACCCCGTGGCGATTACTGCGCTCTGG 135  
 QY 170 CCGAGATCACTGCGCGCAACGTTGTCTAGTTGCACTGCTGGGCC 216  
 Db 134 ATCAATTCATCCGGCCAAAGTTGGCCGATCTCACTGCGCTGGACC 88

RESULT 13  
 BH400384/c

LOCUS BH400384 815 bp DNA linear GSS 11-DEC-2001  
 DEFINITION AG-ND-160D21.TR ND-TAM Anopheles gambiae genomic clone  
 AG-ND-160D21, genomic survey sequence.

ACCESSION BH400384  
 VERSION BH400384.1 GI:17346600  
 KEYWORDS GSS.  
 SOURCE Anopheles gambiae (African malaria mosquito)  
 ORGANISM Anopheles gambiae

REFERENCE 1 (bases 1 to 815)  
 AUTHORS Hong Y.S., Hogan J.R., Wang X., Sarkar A., Sim C., Loftus B.J.,  
 Ren C., Huff B.R., Carlile J.L., Black K., Zhang H.-B.,  
 Gardner M.J. and Collins F.H.

TITLE Construction of a BAC library and generation of BAC end  
 sequence-tagged connectors for genome sequencing of the African  
 malaria mosquito Anopheles gambiae  
 JOURNAL Mol. Genet. Genomics 268 (6), 720-728 (2003)

MEDLINE 22542063  
 PUBMED 12655398

COMMENT Other\_GSSs: AG-ND-160D21.TF  
 Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: bjoftus@igr.org

FEATURES  
 source  
 1..815  
 /organism="Anopheles gambiae"  
 /mol\_type="genomic DNA"  
 /strain="PEST"  
 /db\_xref="taxon:7165"  
 /clone.lib="AG-ND-160D21"  
 /clone.lib="ND-TAM"  
 /note="Vector: pCBAC1; Site\_1: HindIII"

## ORIGIN

Query Match 2.7%; Score 46.2; DB 8; Length 815;  
 Best Local Similarity 64.5%; Pred. No. 0.7;  
 Matches 69; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 110 CCGCTGTGAATGATTAATTAATGAGCCGCAACCAAGAAACTATGCCACTGCCCTCGA 169  
 Db 191 CCCCCGCGATTTGAGAACTAGGAGCAACCCCGTGGCGATTACTGCGCTCTGG 132  
 QY 170 CCGAGATCACTGCGCGCAACGTTGTCTAGTTGCACTGCTGGGCC 216  
 Db 131 ATCAATTCATCCGGCCAAAGTTGGCCGATCTCACTGCGCTGGACC 85

## RESULT 14

LOCUS CA107253 639 bp mRNA linear EST 23-SEP-2003  
 DEFINITION SCRFHR1007G10.g HR1 Saccharum officinarum cDNA clone SCRFHR1007G10  
 5' mRNA sequence.

ACCESSION CA107253  
 VERSION CA107253.1 GI:34960560  
 KEYWORDS EST.  
 SOURCE Saccharum officinarum

## ORGANISM

Saccharum officinarum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD  
clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum  
complex.

REFERENCE  
1 (bases 1 to 639)  
Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.

AUTHORS  
The libraries that made SUCEST

JOURNAL  
Genet. Mol. Biol. 24 (1-4), 1-7 (2001)

COMMENT  
Contact: Arruda P

Centro de Biologia Molecular e Engenharia Genetica  
Universidade Estadual de Campinas  
Caixa Postal 6010, 13083-970, Campinas SP, Brazil

Tel: 55 19 3788 1137  
Fax: 55 19 3788 1089

Email: paruda@unicamp.br

Clone distribution: clone distribution information can be found  
through the Brazilian Clone Collection Center (BCCC) at  
http://www.bcccenter.fcav.unesp.br

Plate: 007 row: G column: 10  
Seq primer: T7 Promoter Primer.

## FEATURES

source  
1. 639  
Location/Qualifiers

/organism="Saccharum officinarum"  
/mol\_type="rRNA"  
/db\_xref="taxon:4547"  
/clone="SCRPHR1007G10"  
/lab\_host="DH10B"  
/clone\_lib="HR1"  
/note="Organ: seedlings inoculated with Herbaspirillum  
rubrisubalbicans; Vector: pSPori1; Site\_1: SalI; Site\_2:  
NotI; An unidirectional cDNA library generated from  
[seedlings inoculated with Herbaspirillum  
rubrisubalbicans]. cDNA was prepared from poly(A)<sup>+</sup> mRNA  
using Superscript Plasmid System Kit (Invitrogen). The  
double-strand cDNAs were fractionated in a sepharose  
CL-2B 40cm-columns and fragments sizing between 0.8 and  
1.5 kb were directionally cloned into the vector. Details  
of each source of RNA and library construction can be  
obtained at http://sucest.fad.ic.unicamp.br/public"

## ORIGIN

Query Match 2.6%; Score 46; DB 6; Length 639;

Best Local Similarity 50.4%; Pred. No. 0.77;  
Matches 138; Conservative 0; Mismatches 135; Indels 1; Gaps 1;

814 ACCGCCGCGGCA CGCTGTATGACCAACACCCGCTTGGCGTCCGACACGGGC 873  
192 ACCACGCGACGGCGCGCGGACGACGCGCGGTGACTCGGCTTCTG-CGTGGCC 250  
874 GAAATTCTGCGGTGACCAACCGCTGCGCGGACCACTGGGACCAAGATGACGTTTC 933  
251 GAGCTGGGCAATGACCGGAGAGCCCGGACGCTGACACCTTGGGCTTGGCAAGGTGGCC 310  
934 GAGATGATGCTCCCAACGTCGATGTGCAACCTCGCCGAGATGAGAGGCTTGGCGGCC 993  
311 GCGGTGACGGGCGTCAACACGCGGACCGCGCTGTACGACATCAAGCCCTGCTCGGCC 370  
994 ATCAACCCCAATGCGCGGACGCGCGCGCTGTGTCTGACGCGTTCGCTTGCAGAAC 1053  
371 ATGAGGCTTGGTGGCGGTGGCGGCGGATGCGCCGACGCGCGGCGCTTGAAGAAGTGC 430  
1054 GGCACGATGTGTCGTTGATGCGGCGCTCGGGCG 1087  
431 GGAAGCTGTACGACTCCGTGGGTTGCGGTTTCG 464

## RESULT 15

LOCUS

CL659599 709 bp DNA linear GSS 09-JUL-2004

DEFINITION  
PRI0134C.F01 - PRI0134C.B21 (709) Mixed stage fosmid library of P.  
pacificus var. California Pristionchus pacificus genomic, genomic  
survey sequence.

## ACCESSION

CL659599 GI:50143867

## VERSION

GSS.

## KEYWORDS

Pristionchus pacificus

## SOURCE

Pristionchus pacificus

## ORGANISM

Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; Pristionchus.

## REFERENCE

1 (bases 1 to 709)  
Srinivasan,J., Otto,G.W., Kahlw,U., Geisler,R. and Sommer,R.J.

## AUTHORS

Appad: an AcedB database for the nematode satellite organism  
Pristionchus pacificus

Nucleic Acids Res. 32 (1), D421-D422 (2004)

## JOURNAL

Contact: Sommer R

## COMMENT

Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498

Email: ralf.sommer@uebingen.mpg.de

This library was generated at Caltech, Pasadena, USA and end  
sequenced at Vancouver, Canada.

Seq primer: T7  
Class: fosmid ends.

## FEATURES

source  
1. 709  
Location/Qualifiers

/organism="Pristionchus pacificus"  
/mol\_type="genomic DNA"  
/strain="California"  
/db\_xref="taxon:54126"  
/clone\_lib="Mixed stage fosmid library of P. pacificus  
var. California"  
/note="Vector: pEplfos-5 Fosmid vector"

## ORIGIN

Query Match 2.6%; Score 46; DB 9; Length 709;

Best Local Similarity 53.3%; Pred. No. 0.78; 85; Indels 0; Gaps 0;  
Matches 97; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

1464 GCCCGTTTTCAGACGCGGCGGTGCTTCAACGCGCGGACCGGCTATTTCG 1523  
369 GCCAATCTCCACGGCGGTAAAGTCTGTTATCGCGCTACGCGAGATTAACCTGCG 428  
1524 TGCCCTGACGCGGAAACCGGCGAGACTTTGTGCGAGCCCGTTCGACGCTCGCGAC 1583  
429 CGTTTACAAATGACGACGCGGTGAAACCTGTGCGCGGCTGTTTACACGCGGCTGCTCA 488  
1584 GGGGACGCGATACGCTACGAGTTGACGCGCGTGCATATATGACCATCGTGGGCGG 1643  
489 GGTACCGCAATGACTTATGAAGTAAAGTAAAGTATGTGTATCTCCGACGGCG 548  
1644 TC 1645  
549 TC 550

Search completed: November 22, 2004, 18:28:03  
Job time : 5789 secs